GenCore version 5.1.9
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2006, 21:41:11 ; Search time 7437 Seconds
(without alignments)
10705.196 Million cell updates/sec

Title: US-10-680-860A-13
Perfect score: 1245
Sequence: 1 gtgctccagcgcgtcgatct.....aacccaccccgagaggatga 1245

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_env:*
2: gb_pat:*
3: gb_ph:*
4: gb_pl:*
5: gb_pr:*
6: gb_ro:*
7: gb_sts:*
8: gb_sy:*
9: gb_un:*
10: gb_vi:*
11: gb_ov:*
12: gb_htg:*
13: gb_in:*
14: gb_om:*
15: gb_ba:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1245	100.0	1245	2	CQ801152	CQ801152 Sequence
	2	1245	100.0	11171	2	CQ801141	CQ801141 Sequence
	3	794	63.8	2381	2	E06690	E06690 DNA encodin
	4	794	63.8	2381	15	STMMDMBC	M93958 Streptomyce
	5	628.4	50.5	1810	2	E09625	E09625 Streptomyce
	6	628.4	50.5	3267	15	STMACYA	D30759 Streptomyce
	7	356.8	28.7	1851	2	I08189	I08189 Sequence 1
c	8	355.4	28.5	2749	2	E02280	E02280 DNA encodin
c	9	355.4	28.5	2749	2	AR363942	AR363942 Sequence
c	10	355.4	28.5	2749	15	STMACYB	D31821 Streptomyce
	11	351.2	28.2	1167	2	AR363943	AR363943 Sequence
	12	345	27.7	1646	2	E28865	E28865 DNA contain
	13	317	25.5	1592	15	STMMPT	D63662 Streptomyce
	14	246.6	19.8	48200	15	AY735112	AY735112 Streptomy
	15	219.2	17.6	210614	15	AB088224	AB088224 Streptomy
c	16	215.6	17.3	110000	15	BA000030_46	Continuation (47 o
	17	192.8	15.5	47981	2	AX112026	AX112026 Sequence
	18	192.8	15.5	47981	15	AF263245	AF263245 Micromono
	19	184	14.8	1137	2	DD164728	DD164728 Biosynthe
c	20	184	14.8	82746	2	DD164760	DD164760 Biosynthe
c	21	184	14.8	82746	15	AF453501	AF453501 Actinosyn
c	22	165.2	13.3	42074	15	SGR578458	AJ578458 Streptomy
	23	98.4	7.9	110000	15	AP006618_16	Continuation (17 o
	24	86.6	7.0	1826	2	AR078690	AR078690 Sequence
	25	86.6	7.0	1826	2	AR217220	AR217220 Sequence
	26	86.6	7.0	1826	2	AR565625	AR565625 Sequence
c	27	85.6	6.9	110000	15	CP000251_06	Continuation (7 of
	28	83.2	6.7	110000	15	CP000251_20	Continuation (21 o
	29	82.4	6.6	3957	2	A45258	A45258 Sequence 2
	30	82.4	6.6	3957	2	AR606608	AR606608 Sequence
	31	82.4	6.6	110000	15	CP000251_32	Continuation (33 o
	32	82.4	6.6	154746	2	CS208162	CS208162 Sequence
c	33	82.4	6.6	154746	2	CS208162	CS208162 Sequence
	34	82.4	6.6	154746	2	AR607375	AR607375 Sequence
c	35	82.4	6.6	154746	2	AR607375	AR607375 Sequence
	36	82.4	6.6	154746	10	HSV2HG52	Z86099 Herpes simp
c	37	82.4	6.6	154746	10	HSV2HG52	Z86099 Herpes simp
	38	81.6	6.6	110000	15	CP000251_35	Continuation (36 o
c	39	81.2	6.5	125020	5	AF429315	AF429315 Homo sapi
	40	80.8	6.5	110000	15	BA000035_15	Continuation (16 o
c	41	80.4	6.5	80472	12	AC174340	AC174340 Medicago
	42	80.2	6.4	150715	10	AY714813	AY714813 Cercopith
c	43	80.2	6.4	150715	10	AY714813	AY714813 Cercopith
c	44	78.6	6.3	136753	12	AC149968	AC149968 Strongylo
c	45	78.4	6.3	110000	15	CP000251_27	Continuation (28 o

ALIGNMENTS

RESULT 1
CQ801152
LOCUS CQ801152 1245 bp DNA linear PAT 05-MAY-2004
DEFINITION Sequence 13 from Patent WO2004033689.
ACCESSION CQ801152
VERSION CQ801152.1 GI:47057924
KEYWORDS .
SOURCE Streptomyces ambofaciens
ORGANISM Streptomyces ambofaciens
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Blondelet-Rouault,M.H., Dominguez,H., Darbon-Rongere,E.,
Gerbaud,C., Gondran,A., Karray,F., Lacroix,P., oestreicher-mermet-
Bouvier,N., Pernodet,J.L. and Tuphile,K.
TITLE Polypeptides involved in spiramycin biosynthesis, nucleotide
sequences encoding said polypeptides and uses thereof
JOURNAL Patent: WO 2004033689-A 13 22-APR-2004;
Aventis Pharma S.A. (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES Location/Qualifiers
source 1. .1245
/organism="Streptomyces ambofaciens"
/mol_type="unassigned DNA"
/db_xref="taxon:1889"
CDS 1. .1245
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAG28207.1"
/db_xref="GI:47057925"
/translation="MLQRVDLSSLTGLRWYAALTVFACHIAQQGFFADQQVGSALLHI
TPLGSMAVSIFFILSGFVLAWSARDEDSVPTFWRRRIAKIYPLHLATFGIAALIIFSL
SEPVLPGGSVWDGLVPNVLLVQSWLPDATLTASFNTPSWSLSCEIAFYLSFPLWYRLV
RRIPARRLWWCAAGIAVAVTCVPLLAGLLPASEEVAPGMSLNEVWFAYWLPPVRMLEF
VLGIVMALILRAGIWKGPGPAVCTALLAASYGLTQMVPPIFTLVACSVVPAALLITAL
ADADVHGRRTGLRSATLVRLGQWSFAFYLVHPLIIRYGHRLMGGDLGYERQWSTPAAI
ALSLGMLGVAVLAGGLLHTVVEQPCMRLFGSRRSASRPKPGATAAPRNSPAADAAGVP
LLPGVPGPAHTPAATNEPTPRG"
ORIGIN

```
  Query Match           100.0%;  Score 1245;  DB 2;  Length 1245;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-284;
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60

Qy         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120

Qy        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180

Qy        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240

Qy        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300

Qy        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360

Qy        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420

Qy        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480

Qy        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540

Qy        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600

Qy        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660

Qy        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720

Qy        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780

Qy        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840

Qy        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900

Qy        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960

Qy        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020

Qy       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080

Qy       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140

Qy       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200

Qy       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245


RESULT 2
CQ801141
LOCUS       CQ801141               11171 bp    DNA     linear   PAT 05-MAY-2004
DEFINITION  Sequence 2 from Patent WO2004033689.
ACCESSION   CQ801141
VERSION     CQ801141.1  GI:47057913
KEYWORDS    .
SOURCE      Streptomyces ambofaciens
  ORGANISM  Streptomyces ambofaciens
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Blondelet-Rouault,M.H., Dominguez,H., Darbon-Rongere,E.,
            Gerbaud,C., Gondran,A., Karray,F., Lacroix,P., oestreicher-mermet-
            Bouvier,N., Pernodet,J.L. and Tuphile,K.
  TITLE     Polypeptides involved in spiramycin biosynthesis, nucleotide
            sequences encoding said polypeptides and uses thereof
  JOURNAL   Patent: WO 2004033689-A 2 22-APR-2004;
            Aventis Pharma S.A. (FR); CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .11171
                     /organism="Streptomyces ambofaciens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:1889"
ORIGIN

  Query Match           100.0%;  Score 1245;  DB 2;  Length 11171;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-284;
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 4120

Qy         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 4180

Qy        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 4240

Qy        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 4300

Qy        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 4360

Qy        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 4420

Qy        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4421 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 4480

Qy        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 4540

Qy        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 4600

Qy        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 4660

Qy        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 4720

Qy        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 4780

Qy        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 4840

Qy        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 4900

Qy        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 4960

Qy        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 5020

Qy        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 5080

Qy       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5081 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 5140

Qy       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5141 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 5200

Qy       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5201 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 5260

Qy       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       5261 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 5305


RESULT 3
E06690
LOCUS       E06690                  2381 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA encoding Streptomyces 3'-acylase.
ACCESSION   E06690
VERSION     E06690.1  GI:2174872
KEYWORDS    JP 1994038750-A/1.
SOURCE      Streptomyces mycarofaciens
  ORGANISM  Streptomyces mycarofaciens
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 2381)
  AUTHORS   Hara,O., Nagaoka,K. and Shii,R.H.
  TITLE     ENZYME FOR ACYLATING 3-POSITION OF MACROLIDE ANTIBIOTIC AND GENE
            CODING THE SAME
  JOURNAL   Patent: JP 1994038750-A 1 15-FEB-1994;
            MEIJI SEIKA KAISHA LTD
COMMENT     OS   Streptomyces mycarofaciens
            PN   JP 1994038750-A/1
            PD   15-FEB-1994
            PF   09-AUG-1991 JP 1991224814
            PI   HARA OSAMU, NAGAOKA KOZO, SHII RICHIYAADO HATSUCHINSON PC
            C12N9/78,C12N1/21,C12N15/55,C12P17/08,(C12N1/21,C12R1:465), PC
            (C12P17/08,
            PC   C12R1:465);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=SF-837;
            CC   *source: clone=pOH23;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   mat_peptide     126. .1286
            FT                   /product='Streptomyces 3'-acylase'.
FEATURES              Location/Qualifiers
     source           1. .2381
                      /organism="Streptomyces mycarofaciens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:1949"
ORIGIN

  Query Match             63.8%;  Score 794;  DB 2;  Length 2381;
  Best Local Similarity   80.9%;  Pred. No. 1.8e-177;
  Matches  937;  Conservative    0;  Mismatches  220;  Indels    1;  Gaps    1;

Qy         12 CGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTG 71
              ||||   ||| |||| |||||||||||| |||||||  |||||  |  ||||||||||||
Db        140 CGTCCGCCTGCCGTCGCTCACCGGCCTTCGCTGGTTCGCGGCCTTAGCGGTATTCGCCTG 199

Qy         72 TCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACAT 131
              ||||||  |||||||||    |||||||| ||||||||||| || | ||| ||||| |||||
Db        200 TCACATTGCCCAGCAGCAATTCTTCGCTGACCAGCAGGTCGGAACCGCGCTGCTACACAT 259

Qy        132 CACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGC 191
              |||| |||||||||| || |||||||| |||||||| | || || || || || || || ||
Db        260 CACCACGCTCGGTTCGATCGCGGTCTCGGTCTTCTTCCTGCTCAGCGGTTTTGTTCTGGC 319

Qy        192 CTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGAT 251
              |||||||||||| ||| |||||||  ||  ||||||||||||||||||  | || |||||
Db        320 GTGGTCGGCCCGTGACAAGGACTCTGTCACGACTTTCTGGCGGCGCCGATTTGCCAAGAT 379

Qy        252 CTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGA 311
              ||| |||||||| |||| || |||  ||| ||||   |||||||||||||||  ||||
Db        380 CTACCCGCTGCACCTCGTAACTTTCCTCATAGCGGGCGTCATCATTTTCTCCCTCGCGGA 439

Qy        312 GCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCA 371
              ||||   || ||||| |||||||||||||||| ||||| ||| || ||||||| |||||
Db        440 GCCGACTCTGCCCGGTGGTTCCGTATGGGACGGTCTGGTTCCCGATCTTCTGCTGGTGCA 499

Qy        372 GTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTC 431
              |||||||||||||  | ||  | |  ||| | |||||||| || || ||||||||||||
Db        500 GTCCTGGCTTCCCGAACCCACGATTATCGCCGGGTTCAACACCCCGAGTTGGTCGCTCTC 559

Qy        432 CTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCC 491
              ||||||  |||| ||||||||| |||||||||||||||| |||||||||| | ||||||
Db        560 CTGTGAATTCGCTTTCTATCTGACGTTCCCGCTGTGGTATCGGCTGGTGCGGAAGATTCC 619

Qy        492 CGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCT 551
              | |||||||||||||||||| ||||| || || |||| ||||||||  ||||| ||  |
Db        620 GGTACGGCGGCTGTGGTGGTGTGCCGCCGGTATTGCCGCGGCCGTGATTTGTGTACCGTT 679

Qy        552 GCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGT 611
              || || ||| | ||||||||||||  |||  |||||||||||||| |||||||||| |
Db        680 CGTGACGAGCCAGTTCCCGGCGAGCGCCGAGACGGCCCCCGGGATGCCGCTCAACGAGCT 739
```

```
Qy        612 CTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGAT 671
              |||||||| | |||||| |||||||||| ||||||||||||||||||||||||||||||
Db        740 GTGGTTCGCCTGCTGGCTGCCGCCGGTGCGGATGCTGGAGTTCGTCCTCGGCATCGTGAT 799

Qy        672 GGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCT 731
              |||||||||||||||| |||| | ||| |||||||||  || |  || | ||  |||
Db        800 GGCGCTGATCCTGCGCACGGGCGTGTGGCGGGGCCCCGGAGTGGTGTCCTCCGCGCTGCT 859

Qy        732 CCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTG 791
              |||||||||   |||||| |||||||| |||||||||||| |||||| ||| ||| ||
Db        860 GCTCGCCGCGGCCTACGGCGTCACCCAGGTGGTGCCCCCGATGTTCACCATCGCCGCGTG 919

Qy        792 CTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCG 851
              |||| |||| || ||||| |||||||||| ||  ||||| | ||||| ||||| ||||
Db        920 CTCCATCGTGCCCGCCGCCCTGCTGATCACCGCTTTGGCCAATGCCGATGTGCAGGGCCT 979

Qy        852 GCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTA 911
              |||||| ||  |||| |||||   ||||||||||||||| ||||||||||||||||||
Db        980 GCGCACCGGTTTGCGCTCGGCGGTACTGGTGCGGCTGGGCGAGTGGTCCTTCGCCTTCTA 1039

Qy        912 CCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTA 971
              ||||| |||||| || |||||||||||| |||||||||||||||||||| ||||||||
Db       1040 TCTGGTTCACTTCATGGTCATCCGCTACGGGCACCGGCTGATGGGCGGCGAACTGGGCTA 1099

Qy        972 CGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGC 1031
              || ||| |||||||||||| ||  ||||   |||||| |||||| |||||||| ||| ||
Db       1100 CGCGCGCCAGTGGAGCACCGCGAGCGCGGGAGCGCTGGCCCTGGCGATGCTGGCGGTCGC 1159

Qy       1032 GGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGG 1091
              | || ||||||| || |||||||||||||||||| ||  | ||||||||||| ||| ||||
Db       1160 GATCGTGGCCGGTGGGCTGCTGCACACCGTCGTGGAGAACCCCTGCATGCGTCTGCTCGG 1219

Qy       1092 CAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG-CGGCTCCCCGGAACTCAC 1150
              | ||||| |  | | | ||   |      |  ||| |||| ||   | ||| |||||||
Db       1220 CCGCCGCCGCCCGGTCGCCACCGCACCAGACCCCGCAACCGACGAGGCTCCGAAACTCAC 1279

Qy       1151 CCGCGGCCGACGCGGCCG 1168
              ||||| | ||||   |||
Db       1280 CCGCGCCTGACGTCACCG 1297


RESULT 4
STMMDMBC
LOCUS       STMMDMBC                2381 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  Streptomyces mycarofaciens acyltransferase (mdmB) and
            O-methyltransferase (mdmC) genes, complete cds.
ACCESSION   M93958
VERSION     M93958.1  GI:153348
KEYWORDS    O-methyltransferase; macrolide 3-O-acetyltransferase.
SOURCE      Streptomyces mycarofaciens
  ORGANISM  Streptomyces mycarofaciens
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 2381)
  AUTHORS   Hara,O. and Hutchinson,C.R.
  TITLE     A macrolide 3-O-acyltransferase gene from the midecamycin-producing
            species Streptomyces mycarofaciens
  JOURNAL   J. Bacteriol. 174 (15), 5141-5144 (1992)
   PUBMED   1629172
COMMENT     Original source text: Streptomyces mycarofaciens (library: ATCC
            21454) DNA.
FEATURES             Location/Qualifiers
     source          1. .2381
                     /organism="Streptomyces mycarofaciens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1949"
                     /tissue_lib="ATCC 21454"
     gene            126. .1289
                     /gene="mdmB"
     CDS             126. .1289
                     /gene="mdmB"
                     /function="catalyzes the addition of acyl groups to the 3
                     position of 16-membered macrolide antibiotics"
                     /codon_start=1
                     /transl_table=11
                     /product="acyltransferase"
                     /protein_id="AAA26781.1"
                     /db_xref="GI:153349"
                     /translation="MPPRVVRLPSLTGLRWFAALAVFACHIAQQQFFADQQVGTALLH
                     ITTLGSIAVSVFFLLSGFVLAWSARDKDSVTTFWRRRFAKIYPLHLVTFLIAGVIIFS
                     LAEPTLPGGSVWDGLVPDLLLVQSWLPEPTIIAGFNTPSWSLSCEFAFYLTFPLWYRL
                     VRKIPVRRLWWCAAGIAAAVICVPFVTSQFPASAETAPGMPLNELWFACWLPPVRMLE
                     FVLGIVMALILRTGVWRGPGVVSSALLLAAAYGVTQVVPPMFTIAACSIVPAALLITA
                     LANADVQGLRTGLRSAVLVRLGEWSFAFYLVHFMVIRYGHRLMGGELGYARQWSTASA
                     GALALAMLAVAIVAGGLLHTVVENPCMRLLGRRRPVATAPDPATDEAPKLTRA"
     gene            1327. .1992
                     /gene="mdmC"
     CDS             1327. .1992
                     /gene="mdmC"
                     /function="biosynthesis of 16-membered macrolide
                     antibiotics"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="O-methyltransferase"
                     /protein_id="AAA26782.1"
                     /db_xref="GI:153350"
                     /translation="MADQTTLSPALLDYARSVALREDGLLRELHDMTAQLPGGRAMQI
                     MPEEAQFLGLLIRLVGARRVLEIGTFTGYSTLCMARALPAGGRIVTCDISDKWPGIGA
                     PFWQRAGVDGLIDLRIGDAARTLAELRERDGDGAFDLVFVDADKAGYLHYYEQALALV
                     RPGGLVAIDNTLFFGRVADPAADDPDTVAVRTLNDLLRDDERVDIALLTVADGITLAR
                     RRE"
ORIGIN

  Query Match           63.8%;  Score 794;  DB 15;  Length 2381;
  Best Local Similarity  80.9%;  Pred. No. 1.8e-177;
  Matches  937;  Conservative    0;  Mismatches  220;  Indels    1;  Gaps    1;

Qy         12 CGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTG 71
              ||||   ||| |||| |||||||||||| |||||||  |||||  |  ||||||||||||
Db        140 CGTCCGCCTGCCGTCGCTCACCGGCCTTCGCTGGTTCGCGGCCTTAGCGGTATTCGCCTG 199

Qy         72 TCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACAT 131
              |||||| |||||||||   |||||||| |||||||||||| || | ||| ||||| |||||
Db        200 TCACATTGCCCAGCAGCAATTCTTCGCTGACCAGCAGGTCGGAACCGCGCTGCTACACAT 259

Qy        132 CACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGC 191
              |||| |||||||||| || ||||||||| |||||||| | || || || || || || ||
Db        260 CACCACGCTCGGTTCGATCGCGGTCTCGGTCTTCTTCCTGCTCAGCGGTTTTGTTCTGGC 319

Qy        192 CTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGAT 251
              |||||||||| ||| ||||||| ||  ||||||||||||||||||  | || |||||
Db        320 GTGGTCGGCCCGTGACAAGGACTCTGTCACGACTTTCTGGCGGCGCCGATTTGCCAAGAT 379

Qy        252 CTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGA 311
              ||| |||||||| |||   || |||  ||| ||||   ||||||||||||||||  ||||
Db        380 CTACCCGCTGCACCTCGTAACTTTCCTCATAGCGGGCGTCATCATTTTCTCCCTCGCGGA 439

Qy        312 GCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCA 371
              ||||   || ||||| |||||||||||||||| ||||| ||| || ||||||| |||||
Db        440 GCCGACTCTGCCCGGTGGTTCCGTATGGGACGGTCTGGTTCCCGATCTTCTGCTGGTGCA 499

Qy        372 GTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTC 431
              |||||||||||||| | ||  | |  ||| | |||||||| || || ||||||||||
Db        500 GTCCTGGCTTCCCGAACCCACGATTATCGCCGGGTTCAACACCCCGAGTTGGTCGCTCTC 559

Qy        432 CTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCC 491
              |||||  |||| |||||||||| |||||||||||||||| ||||||||||| | ||||||
Db        560 CTGTGAATTCGCTTTCTATCTGACGTTCCCGCTGTGGTATCGGCTGGTGCGGAAGATTCC 619

Qy        492 CGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCT 551
              | |||||||||||||||||| ||||| || || |||| |||||||  ||||| ||  |
Db        620 GGTACGGCGGCTGTGGTGGTGTGCCGCCGGTATTGCCGCGGCCGTGATTTGTGTACCGTT 679

Qy        552 GCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGT 611
              || || ||| | |||||||||||  |||  |||||||||||| |||||||||| |
```

```
Db        680 CGTGACGAGCCAGTTCCCGGCGAGCGCCGAGACGGCCCCCGGGATGCCGCTCAACGAGCT 739

Qy        612 CTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGAT 671

Db        740 GTGGTTCGCCTGCTGGCTGCCGCCGGTGCGGATGCTGGAGTTCGTCCTCGGCATCGTGAT 799

Qy        672 GGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCT 731

Db        800 GGCGCTGATCCTGCGCACGGGCGTGTGGCGGGGCCCCGGAGTGGTGTCCTCCGCGCTGCT 859

Qy        732 CCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTG 791

Db        860 GCTCGCCGCGGCCTACGGCGTCACCCAGGTGGTGCCCCCGATGTTCACCATCGCCGCGTG 919

Qy        792 CTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCG 851

Db        920 CTCCATCGTGCCCGCCGCCCTGCTGATCACCGCTTTGGCCAATGCCGATGTGCAGGGCCT 979

Qy        852 GCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTA 911

Db        980 GCGCACCGGTTTGCGCTCGGCGGTACTGGTGCGGCTGGGCGAGTGGTCCTTCGCCTTCTA 1039

Qy        912 CCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTA 971

Db       1040 TCTGGTTCACTTCATGGTCATCCGCTACGGGCACCGGCTGATGGGCGGCGAACTGGGCTA 1099

Qy        972 CGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGC 1031

Db       1100 CGCGCGCCAGTGGAGCACCGCGAGCGCGGGAGCGCTGGCCCTGGCGATGCTGGCGGTCGC 1159

Qy       1032 GGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGG 1091

Db       1160 GATCGTGGCCGGTGGGCTGCTGCACACCGTCGTGGAGAACCCCTGCATGCGTCTGCTCGG 1219

Qy       1092 CAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG-CGGCTCCCCGGAACTCAC 1150

Db       1220 CCGCCGCCGCCCGGTCGCCACCGCACCAGACCCCGCAACCGACGAGGCTCCGAAACTCAC 1279

Qy       1151 CCGCGGCCGACGCGGCCG 1168

Db       1280 CCGCGCCTGACGTCACCG 1297


RESULT 5
E09625
LOCUS       E09625                  1810 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  Streptomyces acyA gene.
ACCESSION   E09625
VERSION     E09625.1  GI:22026252
KEYWORDS    JP 1995177892-A/1.
SOURCE      Streptomyces sp.
  ORGANISM  Streptomyces sp.
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 1810)
  AUTHORS   Arisawa,A., Kawamura,N., Kojima,I., Okamura,K., Tone,H. and
            Okamoto,R.
  TITLE     GANE CODING 3-ACYLATION ENZYME OF MACROLIDE ANTIBIOTIC SUBSTANCE
  JOURNAL   Patent: JP 1995177892-A 1 18-JUL-1995;
            MERCIAN CORP
COMMENT     OS   Streptomyces thermotolerans
            PN   JP 1995177892-A/1
            PD   18-JUL-1995
            PF   22-FEB-1991 JP 1991048753
            PR   01-JUN-1990 JP 90P   141625
            PI   ARISAWA AKIRA, KAWAMURA NAOTO, KOJIMA IKUO, OKAMURA KAZUHIKO,
            PI   TONE HIROSHI, OKAMOTO ROKURO
            PC   C12N15/54,C12N15/11,C12N15/76,C12P19/62//C12N1/21,C12N9/18, PC
              (C12N15/54,
            PC   C12R1:465),(C12P19/62,C12R1:465),(C12P19/62,C12R1:54), PC
            (C12N1/21,C12R1:465);
            CC   strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1810
            FT                   /organism='Streptomyces thermotolerans' FT
            -35_signal      120. .125
            FT   -10_signal      143. .148
            FT   RBS             194. .198
            FT   RBS             201. .205
            FT   CDS             209. .1378
            FT                   /product='AcyA protein'.
FEATURES             Location/Qualifiers
     source          1. .1810
                     /organism="Streptomyces sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1931"
ORIGIN

  Query Match             50.5%;  Score 628.4;  DB 2;  Length 1810;
  Best Local Similarity   71.4%;  Pred. No. 3.6e-138;
  Matches  827;  Conservative    0;  Mismatches  331;  Indels    0;  Gaps    0;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78

Db        230 CTACCTTCACTGACCGGGCTGCGCTGGTTCGCGGCGCTTTCCGTATTCGTCTGCCATATC 289

Qy         79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCCCG 138

Db        290 GCCCAGCAGGGCATTTTCGCCGACCCGGACGTGGCGTCCGCCCTGGGGCACCTCACGTCG 349

Qy        139 CTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCG 198

Db        350 CTCGGCTCGCTCGCGGTCTCCCTCTTCTTCGTGCTGAGCGGCTATGTGCTGACGTGGTCG 409

Qy        199 GCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCG 258

Db        410 GCCCGGGACGGCGACTCCGTCAGAAGCTTCTGGCAGCGTAGGTTCGCCAAGATCTACCCG 469

Qy        259 CTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCGGTA 318

Db        470 CTGCATTTCGTCACCTTCTGCATTGCGGGCGTCATCATCTTCAGTCTTTCCGAGCCGGTG 529

Qy        319 CTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCCTGG 378

Db        530 CTGCCGGGCGGTTCGACGTGGGACGGCATGGTGCCCAACCTGCTGCTGGTGCATTCATGG 589

Qy        379 CTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGTGAG 438

Db        590 CTGCCGGACGCGTACATCGTCTCGGGATTCAACACACCGAGCTGGTCGCTTTCCTGTGAA 649

Qy        439 ATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGG 498

Db        650 ATGGCCTTCTATCTCACGTTTCCGTTGTGGTACCGGCTGCTGCTTCGGATACGGGTGAGC 709

Qy        499 CGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCG 558

Db        710 CGGTTGTGGTGGTACGCGGCCGCACTGGCGCTGGCCGTGGTGTGCATGCCGTTCGTGGCC 769

Qy        559 GGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTC 618

Db        770 CGGCTGCTGCCGGACAGCGCGGAGGTCGTTCCCGGGATGCCGCTGCGGGACATGTGGTTC 829

Qy        619 GCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTG 678

Db        830 ACGTACTGGTTCCCGCCCGTGCGGATGCTGGAGTTCCTCCTCGGCATCGTGCTGGCGCTG 889

Qy        679 ATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCC 738

Db        890 ATCCGGCGCCAGGGGGCGTGGCGGGGGCCCGGAACGGGCACGGCCGCGCTGCTGCTCGGC 949

Qy        739 GCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTC 798

Db        950 GGCGCGTTCGCGCTCAACCAGGTGGTGCCGCCGATGTTCACCCTCACCGCCACCACCGTC 1009
```

```
Qy        799 GTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACG 858
              || ||    ||| |||||||||| | ||    ||| |||| |||  ||| ||| || || ||
Db       1010 GTCCCCATCGCCCTGCTGATCGCCGCCGCGGCGGACGGCGATCTGCGCGGGCGCCGTACC 1069

Qy        859 GGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTC 918
              || |||||  ||||   ||||||  |||||||| ||||||| | ||||||||||||| ||
Db       1070 GGACTGCGCGCGGCCGTGCTGGTCAGGCTGGGCGAGTGGTCGTACGCCTTCTACCTGATC 1129

Qy        919 CACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGG 978
              |||||||||||||| |||||||| ||||||||| |||||||||| | ||| ||||  |||
Db       1130 CACTTCCTGATCATTCGCTACGGGCACCGGCTGCTGGGCGGCGACCAGGGATACGCCCGG 1189

Qy        979 CAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTG 1038
              ||||||  |||||  || ||| |||   |  |   || | | ||||||||  || ||
Db       1190 CAGTGGGACACCCTCGCGGCGCTCGGCATCACAGCGGCGGTACTGGGGGTCACGATCGCC 1249

Qy       1039 GCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGC 1098
              ||  |||   | ||||||| | |||||||| | |||||| |||  |||| |  || |||||
Db       1250 GCGAGCGCGGTCCTGCACATCTTCGTCGAGCGGCCCTGTATGACCCTGCTGCGCGGCCGC 1309

Qy       1099 AGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCC 1158
                |  |  |      | |||   |||| | ||   |      || |      | ||   ||
Db       1310 CGCCCTCCGCAGGGGCCGGCTCCCGACTCCGGGGGCCGCCCGCACCGGGCTCCGCTGGAA 1369

Qy       1159 GACGCGGCCGGCGTGCCC 1176
                ||     ||||| ||
Db       1370 AGGGCATGACGCGTGGCC 1387
```

RESULT 6
STMACYA
LOCUS STMACYA 3267 bp DNA linear BCT 07-FEB-1999
DEFINITION Streptomyces thermotolerans genes for macrolide-resistance determinant (partial cds), cytochrome P-450 protein (complete cds), macrolide antibiotics 3-O-acyltransferase (complete cds) and carbomycin 4-O-metyltransferase (partial cds).
ACCESSION D30759
VERSION D30759.1 GI:551628
KEYWORDS macrolide antibiotics 3-O-acyltransferase; macrolide-resistance determinant; carbomycin 4-O-metyltransferase; cytochrome P-450 protein.
SOURCE Streptomyces thermotolerans
ORGANISM Streptomyces thermotolerans
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 192)
AUTHORS Schoner,B., Geistlich,M., Rosteck,P. Jr., Rao,R.N., Seno,E., Reynolds,P., Cox,K., Burgett,S. and Hershberger,C.
TITLE Sequence similarity between macrolide-resistance determinants and ATP-binding transport proteins
JOURNAL Gene 115 (1-2), 93-96 (1992)
PUBMED 1612454
REFERENCE 2 (sites)
AUTHORS Arisawa,A., Kawamura,N., Takeda,K., Tsunekawa,H., Okamura,K. and Okamoto,R.
TITLE Cloning of the macrolide antibiotic biosynthesis gene acyA, which encodes 3-O-acyltransferase, from Streptomyces thermotolerans and its use for direct fermentative production of a hybrid macrolide antibiotic
JOURNAL Appl. Environ. Microbiol. 60 (7), 2657-2660 (1994)
PUBMED 8074537
REFERENCE 3 (bases 1 to 3267)
AUTHORS Arisawa,A., Tsunekawa,H., Okamura,K. and Okamoto,R.
TITLE Nucleotide sequence analysis of the carbomycin biosynthetic genes including the 3-O-acyltransferase gene from Streptomyces thermotolerans
JOURNAL Biosci. Biotechnol. Biochem. 59 (4), 582-588 (1995)
PUBMED 7772821
REFERENCE 4 (bases 1 to 3267)
AUTHORS Arisawa,A.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-1994) Akira Arisawa, Mercian Corporation Central Laboratories; 4-9-1 Johnan, Fujisawa-shi, Kanagawa 251, Japan
(E-mail:mercian2@city.fujisawa.kanagawa.jp, Tel:0466-35-1511, Fax:0466-35-1530)

FEATURES Location/Qualifiers

```
     source          1. .3267
                     /organism="Streptomyces thermotolerans"
                     /mol_type="genomic DNA"
                     /strain="TH475"
                     /db_xref="taxon:80858"
                     /clone="p53A"
     gene            1. .192
                     /gene="carA"
     CDS             <1. .192
                     /gene="carA"
                     /codon_start=1
                     /transl_table=11
                     /product="macrolide-resistance determinant"
                     /protein_id="BAA06419.1"
                     /db_xref="GI:551629"
                     /translation="VDLLLLDEPTNHLSPALVEELEEALTGYQGTVVVVTHDRRMRSR
                     FNGAHLTLQDGRVAEFTAA"
     stem_loop       211. .241
     CDS             complement(248. .1483)
                     /note="ORF-A"
                     /codon_start=1
                     /transl_table=11
                     /product="putative cytochrome P-450 protein"
                     /protein_id="BAA06420.1"
                     /db_xref="GI:551630"
                     /translation="MTALNLMDPEVLRDPFGAYAEIRAQAPLVRASYPWGAVQWLATR
                     YSDVKAVLTDPRLVNNPANVPEMHLPHPYEQALGDGGIPDEYVRYLAGSILSQDGPGH
                     LRLRRLSSRAFTVRRVNALRPRVTELAHRLLASLPDRAQDGVIDVLEDFSYPLSIDVI
                     CEIVGIPEEAREQWHTWGSAFYTMDPAVIGPAVRGMADHLHLLIEQRRATPTGDLLTG
                     LVQAEDEQGEPLTDEEIATLVLTFVTAGNETTAHLIGNGVAALLTHSDQLALLRSDRR
                     LLSQAVDELMRWCTPVQVTQPRYATEDLDVGGVTVRKGEQVVAVIGAAGHDPDRFPDP
                     ERFDITRNHRAPHEAHVGFGFGPHYCLGAALAHQETAIALDTLFDRFPSLALAVPPSA
                     LERQPFPGAWRLKSLPVRL"
     RBS             complement(1493. .1498)
     -35_signal      1577. .1582
     -10_signal      1600. .1605
     RBS             1651. .1655
     RBS             1658. .1662
     gene            1666. .2835
                     /gene="acyA"
     CDS             1666. .2835
                     /gene="acyA"
                     /experiment="experimental evidence, no additional details
                     recorded"
                     /note="sequence homologous to other O-acyltransferases"
                     /codon_start=1
                     /transl_table=11
                     /product="macrolide antibiotics 3-O-acyltransferase"
                     /protein_id="BAA06421.1"
                     /db_xref="GI:551631"
                     /translation="MESRVERLPSLTGLRWFAALSVFVCHIAQQGIFADPDVASALGH
                     LTSLGSLAVSLFFVLSGYVLTWSARDGDSVRSFWQRRFAKIYPLHFVTFCIAGVIIFS
                     LSEPVLPGGSTWDGMVPNLLLVHSWLPDAYIVSGFNTPSWSLSCEMAFYLTFPLWYRL
                     LLRIRVSRLWWYAAALALAVVCMPFVARLLPDSAEVVPGMPLRDMWFTYWFPPVRMLE
                     FLLGIVLALIRRQGAWRGPGTGTAALLLGGAFALNQVVPPMFTLTATTVVPIALLIAA
                     AADGDLRGRRTGLRAAVLVRLGEWSYAFYLIHFLIIRYGHRLLGGDQGYARQWDTLAA
                     LGITAAVLGVTIAASAVLHIFVERPCMTLLRGRRPPQGPAPDSGGRPHRAPLERA"
     CDS             2839. .>3267
                     /note="ORF-B"
                     /codon_start=1
                     /transl_table=11
                     /product="putative carbomycin 4-O-metyltransferase"
                     /protein_id="BAA06422.1"
                     /db_xref="GI:551632"
                     /translation="MADQTVLSPALLEYARSVSLRDDAVLRELRELTAALPGGRAMQI
                     MPEEAQLLALLIRLTGAAQVLEIGTFTGYSTLCMARALPPGGRIVTCDITERWPGVGA
                     PFWRQAGVADRIDLRIGDAARTLSELRAHEGDGIFDLVFVD"
```

ORIGIN

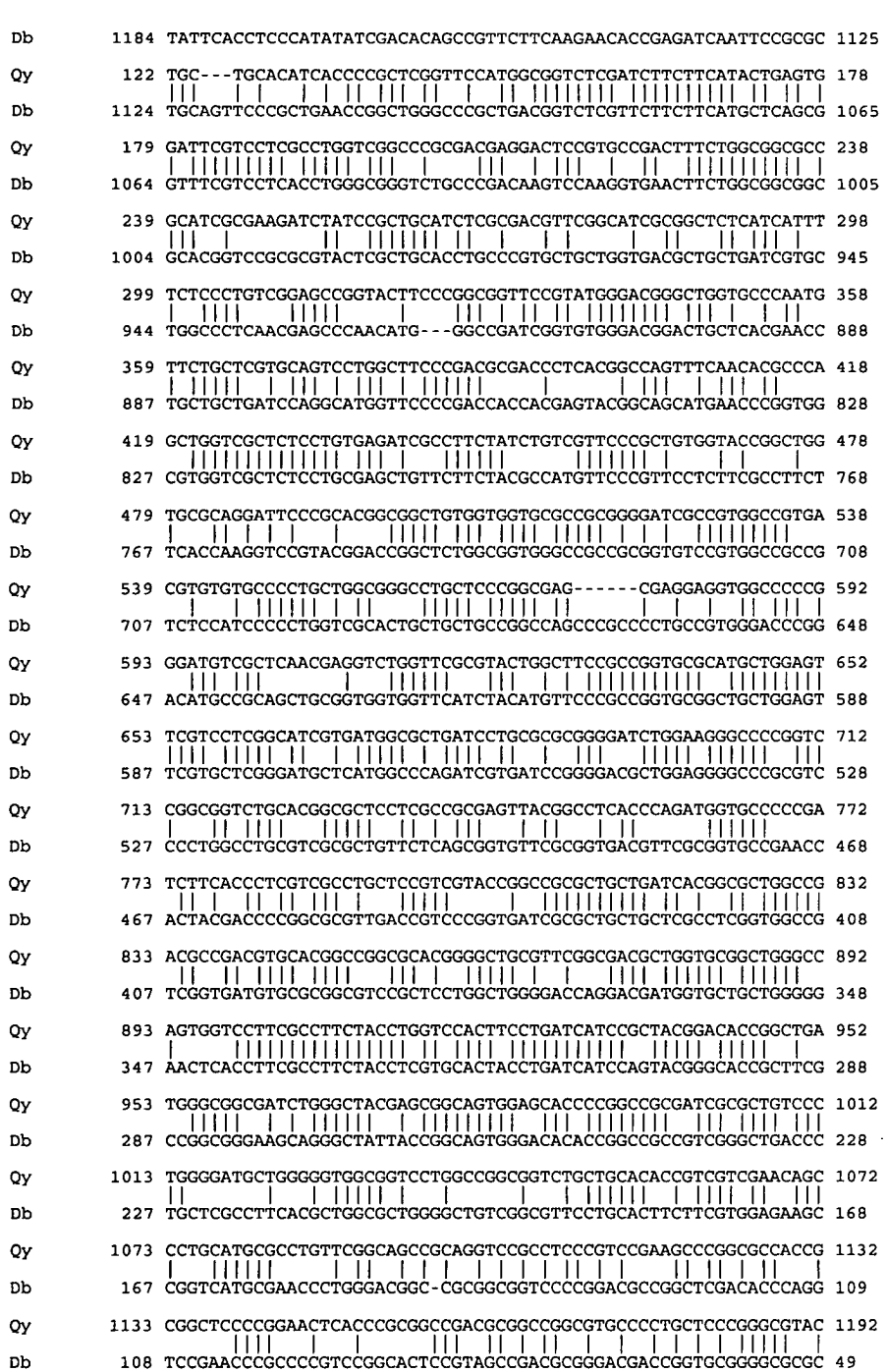
Query Match 50.5%; Score 628.4; DB 15; Length 3267;

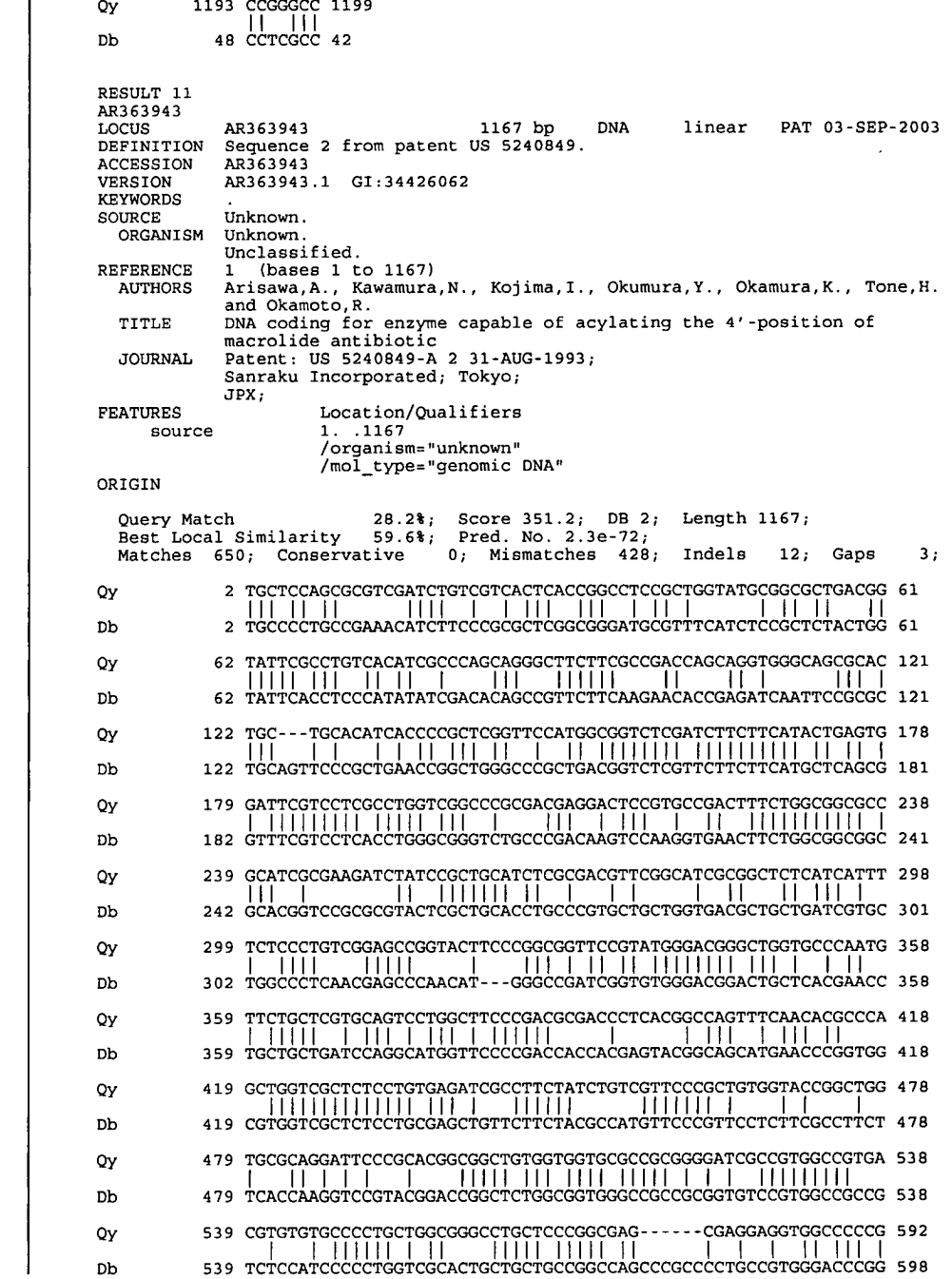
```
  Best Local Similarity   71.4%;  Pred. No. 3.8e-138;
  Matches  827;  Conservative    0;  Mismatches  331;  Indels    0;  Gaps    0;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
              ||  | |||||  |||||  ||  |||||||   ||||||||   | |||||||  ||| || |||
Db       1687 CTACCTTCACTGACCGGGCTGCGCTGGTTCGCGGCGCTTTCCGTATTCGTCTGCCATATC 1746

Qy         79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCCCG 138
              ||||||||||||  | ||||||||||  | | ||||     ||| |||  ||||  ||||  ||
Db       1747 GCCCAGCAGGGCATTTTCGCCGACCCGGACGTGGCGTCCGCCCTGGGGCACCTCACGTCG 1806

Qy        139 CTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCG 198
              |||||  ||   | ||||||||  ||||||||  | |||||  || |   ||  ||   | ||||||
Db       1807 CTCGGCTCGCTCGCGGTCTCCCTCTTCTTCGTGCTGAGCGGCTATGTGCTGACGTGGTCG 1866

Qy        199 GCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCG 258
              |||||  ||||  ||||||||     |  |||||||  |||   |  ||||  |||||||| |||
Db       1867 GCCCGGGACGGCGACTCCGTCAGAAGCTTCTGGCAGCGTAGGTTCGCCAAGATCTACCCG 1926

Qy        259 CTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCGGTA 318
              ||||||  |||   ||  |||  ||||  ||||    |||||||  |||    ||  ||  ||||||||
Db       1927 CTGCATTTCGTCACCTTCTGCATTGCGGGCGTCATCATCTTCAGTCTTTCCGAGCCGGTG 1986

Qy        319 CTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCCTGG 378
              ||  ||  ||||||||     ||||||||  |||||||||   | |||||  |||||  ||  |||
Db       1987 CTGCCGGGCGGTTCGACGTGGGACGGCATGGTGCCCAACCTGCTGCTGGTGCATTCATGG 2046

Qy        379 CTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGTGAG 438
              ||  ||  ||||||   |  ||     |  |  ||||||||  ||  ||||||||||  |||||||
Db       2047 CTGCCGGACGCGTACATCGTCTCGGGATTCAACACACCGAGCTGGTCGCTTTCCTGTGAA 2106

Qy        439 ATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGG 498
              ||  ||||||||||   ||||  |||  |||||||||||||||  |||    ||||  |   |    |
Db       2107 ATGGCCTTCTATCTCACGTTTCCGTTGTGGTACCGGCTGCTGCTTCGGATACGGGTGAGC 2166

Qy        499 CGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCG 558
              |||  |||||||||  |||  ||  |    |  ||  ||||||||   |||   ||||   |  ||||
Db       2167 CGGTTGTGGTGGTACGCGGCCGCACTGGCGCTGGCCGTGGTGTGCATGCCGTTCGTGGCC 2226

Qy        559 GGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTC 618
              |  |||||  ||||   ||||  ||||||  |  |||||||||  ||||     ||  |  ||||||
Db       2227 CGGCTGCTGCCGGACAGCGCGGAGGTCGTTCCCGGGATGCCGCTGCGGGACATGTGGTTC 2286

Qy        619 GCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTG 678
              ||||||||  | |||||  |||||  ||||||||||||  ||||||||||||||||  ||||||||
Db       2287 ACGTACTGGTTCCCGCCCGTGCGGATGCTGGAGTTCCTCCTCGGCATCGTGCTGGCGCTG 2346

Qy        679 ATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCC 738
              ||||  ||||   ||||   |||  |||  |||||   |||       |  ||   |||  ||||  |
Db       2347 ATCCGGCGCCAGGGGGCGTGGCGGGGGCCCGGAACGGGCACGGCCGCGCTGCTGCTCGGC 2406

Qy        739 GCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTC 798
              |      | ||  ||||  ||||  |||||||  |||||  |||||||||   ||||   | |||||
Db       2407 GGCGCGTTCGCGCTCAACCAGGTGGTGCCGCCGATGTTCACCCTCACCGCCACCACCGTC 2466

Qy        799 GTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACG 858
              ||  ||   |||  |||||||||  | ||   |||  ||||  |||   |||  |||  ||  ||  ||
Db       2467 GTCCCCATCGCCCTGCTGATCGCCGCCGCGGCGGACGGCGATCTGCGCGGGCGCCGTACC 2526

Qy        859 GGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTC 918
              ||  |||||   ||||   ||||||  ||||||||  |||||||  |  ||||||||||||  ||
Db       2527 GGACTGCGCGCGGCCGTGCTGGTCAGGCTGGGCGAGTGGTCGTACGCCTTCTACCTGATC 2586

Qy        919 CACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGG 978
              ||||||||||||||  ||||||||  |||||||||  ||||||||||  |  |||  ||||  |||
Db       2587 CACTTCCTGATCATTCGCTACGGGCACCGGCTGCTGGGCGGCGACCAGGGATACGCCCGG 2646

Qy        979 CAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTG 1038
              ||||||  |||||  ||  |||  |||   |  |   ||  |  |  ||||||||  ||  ||
Db       2647 CAGTGGGACACCCTCGCGGCGCTCGGCATCACAGCGGCGGTACTGGGGGTCACGATCGCC 2706

Qy       1039 GCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGC 1098
              ||  |||   | |||||||  | |||||||  | ||||||  |||  ||||  |  ||  |||||
Db       2707 GCGAGCGCGGTCCTGCACATCTTCGTCGAGCGGCCCTGTATGACCCTGCTGCGCGGCCGC 2766

Qy       1099 AGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCC 1158
              |  |  |    | |||   |||| | ||   |      || |      | ||   ||
Db       2767 CGCCCTCCGCAGGGGCCGGCTCCCGACTCCGGGGGCCGCCCGCACCGGGCTCCGCTGGAA 2826

Qy       1159 GACGCGGCCGGCGTGCCC 1176
                ||     ||||| ||
Db       2827 AGGGCATGACGCGTGGCC 2844


RESULT 7
I08189
LOCUS       I08189                 1851 bp    DNA     linear   PAT 02-DEC-1994
DEFINITION  Sequence 1 from Patent EP 0354641.
ACCESSION   I08189
VERSION     I08189.1  GI:589096
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1851)
  AUTHORS   Epp,J.K. and Schoner,B.E.
  TITLE     Recombinant DNA encoding a 4''-0-isovaleryl acylase
  JOURNAL   Patent: EP 0354641-A2 1 14-FEB-1990;
FEATURES             Location/Qualifiers
     source          1. .1851
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           28.7%;  Score 356.8;  DB 2;  Length 1851;
  Best Local Similarity   58.0%;  Pred. No. 1.1e-73;
  Matches  711;  Conservative    0;  Mismatches  502;  Indels   13;  Gaps    4;

Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61
              ||| || ||        ||||  |  | |||  |||  | || |       | || ||   ||
Db        582 TGCCCCTGCCGAAACATCTTCCCGCGCTCGGCGGGATGCGTTTCATCTCCGCTCTACTGG 641

Qy         62 TATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCAC 121
              |||||  |||   || ||  |    |||  ||||||    ||   || |       ||| |
Db        642 TATTCACCTCCCATATATCGACACAGCCGTTCTTCAAGAACACCGAGATCAATTCCGCGC 701

Qy        122 TGC---TGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTG 178
              |||   | |   | | || ||| ||  |  || ||||||||  ||||||||| || || |
Db        702 TGCAGTTCCCGCTGAACCGGCTGGGCCCGCTGACGGTCTCGTTCTTCTTCATGCTCAGCG 761

Qy        179 GATTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCC 238
              | |||||||| |||||  |||  |    |||  | |||  |  ||  ||||||||||| |
Db        762 GTTTCGTCCTCACCTGGGCGGGTCTGCCCGACAAGTCCAAGGTGAACTTCTGGCGGCGGC 821

Qy        239 GCATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTT 298
              |||  |          ||  |||||||  ||  |   | |       |  ||   || ||| |
Db        822 GCACGGTCCGCGCGTACTCGCTGCACCTGCCCGTGCTGCTGGTGACGCTGCTGATCGTGC 881

Qy        299 TCTCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATG 358
              |  ||||     |||||       |    ||| | || || ||||||||  ||| |   | ||
Db        882 TGGCCCTCAACGAGCCCAACAT---GGGCCGATCGGTGTGGGACGGACTGCTCACGAACC 938

Qy        359 TTCTGCTCGTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCA 418
              | |||||  | ||| | ||| | ||||||      |      | |||   | |||  ||
Db        939 TGCTGCTGATCCAGGCATGGTTCCCCGACCACCACGAGTACGGCAGCATGAACCCGGTGG 998

Qy        419 GCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGG 478
              |||||||||||||| ||| |    ||||||       |||||| |    | | |
Db        999 CGTGGTCGCTCTCCTGCGAGCTGTTCTTCTACGCCATGTTCCCGTTCCTCTTCGCCTTCT 1058

Qy        479 TGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGA 538
              |   || | | |    |    ||||| ||| |||| ||||| | | |  |||||||||
Db       1059 TCACCAAGGTCCGTACGGACCGGCTCTGGCGGTGGGCCGCCGCGGTGTCCGTGGCCGCCG 1118

Qy        539 CGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAG------CGAGGAGGTGGCCCCCG 592
                |  | ||||||  | ||    |||||  |||||  ||      |  |  |  || ||| |
```

```
Db       1119 TCTCCATCCCCCTGGTCGCACTGCTGCTGCCGGCCAGCCCGCCCCTGCCGTGGGACCCGG 1178

Qy        593 GGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGT 652

Db       1179 ACATGCCGCAGCTGCGGTGGTGGTTCATCTACATGTTCCCGCCGGTGCGGCTGCTGGAGT 1238

Qy        653 TCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712

Db       1239 TCGTGCTCGGGATGCTCATGGCCCAGATCGTGATCCGGGGACGCTGGAGGGGCCCGCGTC 1298

Qy        713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772

Db       1299 CCCTGGCCTGCGTCGCGCTGTTCTCAGCGGTGTTCGCGGTGACGTTCGCGGTGCCGAACC 1358

Qy        773 TCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCG 832

Db       1359 ACTACGACCCCGGCGCGTTGACCGTCCCGGTGATCGCGCTGCTGCTCGCCTCGGTGGCCG 1418

Qy        833 ACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCC 892

Db       1419 TCGGTGATGTGCGCGGCGTCCCCTCCTGGCTGGGGACCAGGACGATGGTGCTGCTGGGGG 1478

Qy        893 AGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGA 952

Db       1479 AACTCACCTTCGCCTTCTACCTCGTGCACTACCTGATCATCCAGTACGGGCACCGCTTCG 1538

Qy        953 TGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCC 1012

Db       1539 CCGGCGGGAAGCAGGGCTATTACCGGCAGTGGGACACACCGGCCGCCGTCGGGCTGACCC 1598

Qy       1013 TGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGC 1072

Db       1599 TGCTCGCCTTCACGCTGGCGCTGGGGCTGTCGGCGTTCCTGCACTTCTTCGTGGAGAAGC 1658

Qy       1073 CCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG 1132

Db       1659 CGGTCATGCGAACCCTGGGACGGC-CGCGGCGGTCCCCGGACGCCGGCTCGACACCCAGG 1717

Qy       1133 CGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTAC 1192

Db       1718 TCCGAACCCGCCCCGTCCGGCACTCCGTAGCCGACGCGGGACGACCGGTGCGCGGCGCGC 1777

Qy       1193 CCGGGCCCGCGCACACCCCCGCAGCG 1218

Db       1778 CCTCGGGCGCGCCCCGCACCGGTGTG 1803


RESULT 8
E02280/c
LOCUS       E02280                  2749 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA encoding macrolide antibiotics 4'acylation enzyme(acyB).
ACCESSION   E02280
VERSION     E02280.1  GI:2170517
KEYWORDS    JP 1990072880-A/1.
SOURCE      Streptomyces sp.
  ORGANISM  Streptomyces sp.
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 2749)
  AUTHORS   Arisawa,A., Kawamura,N., Kojima,I., Okamura,K., Tone,H. and
            Okamoto,R.
  TITLE     GENE CODING 4-ACYLATION ENZYME OF MACROLIDE ANTIBIOTIC SUBSTANCE
  JOURNAL   Patent: JP 1990072880-A 1 13-MAR-1990;
            SANRAKU INC
COMMENT     OS   Streptomyces thermotolerans
            PN   JP 1990072880-A/1
            PD   13-MAR-1990
            PF   03-MAR-1989 JP 1989050120
            PR   24-MAY-1988 JP 88P   125091
            PI   ARISAWA AKIRA, KAWAMURA NAOTO, KOJIMA IKUO, OKAMURA KAZUHIKO,
            PI   TONE HIROSHI, OKAMOTO ROKURO
            PC   C12N15/55,C12N1/15,C12P19/62//C12N9/00,(C12N15/55,C12R1:465),
            PC   (C12N1/15,
            PC   C12R1:465),(C12N1/15,C12R1:54),(C12P19/62,C12R1:465); CC
            strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=ATCC11416;
            CC   *source: clone=pMAB1;
            FH   Key             Location/Qualifiers
            FH
            FT   mRNA            1. .2769
            FT                   /gene='acyB'
            FT   CDS             Complement(79. .1245)
            FT                   /Product='macrolide antibiotics 4'acylation FT
                              enzyme'
            FT                   /note='acyB1'
            FT   mat_peptide     Complement(76. .1245)
            FT                   /product='macrolide antibiotics 4'acylation
            FT                   enzyme'
            FT   CDS             One-of(1569,1608). .2732
            FT                   /product='macrolide antibiotics 4'acylation
            FT                   enzyme'
            FT                   /note='acyB2'
            FT   mat_peptide     One-of(1569,1608). .2729
            FT                   /product='macrolide antibiotics 4'acylation
            FT                   enzyme'
            FT   3'UTR           Complement(1. .75)
            FT   5'UTR           Complement(1246. .2769)
            FT   5'UTR           1. .one-of(1568,1607)
            FT   3'UTR           2733. .2769.
FEATURES             Location/Qualifiers
     source          1. .2749
                     /organism="Streptomyces sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1931"
ORIGIN

  Query Match           28.5%;  Score 355.4;  DB 2;  Length 2749;
  Best Local Similarity   58.2%;  Pred. No. 2.4e-73;
  Matches  703;  Conservative    0;  Mismatches  491;  Indels   13;  Gaps    4;

Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61

Db       1244 TGCCCCTGCCGAAACATCTTCCCGCGCTCGGCGGGATGCGTTTCATCTCCGCTCTACTGG 1185

Qy         62 TATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCAC 121

Db       1184 TATTCACCTCCCATATATCGACACAGCCGTTCTTCAAGAACACCGAGATCAATTCCGCGC 1125

Qy        122 TGC---TGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTG 178

Db       1124 TGCAGTTCCCGCTGAACCGGCTGGGCCCGCTGACGGTCTCGTTCTTCTTCATGCTCAGCG 1065

Qy        179 GATTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCC 238

Db       1064 GTTTCGTCCTCACCTGGGCGGGTCTGCCCGACAAGTCCAAGGTGAACTTCTGGCGGCGGC 1005

Qy        239 GCATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTT 298

Db       1004 GCACGGTCCGCGCGTACTCGCTGCACCTGCCCGTGCTGCTGGTGACGCTGCTGATCGTGC 945

Qy        299 TCTCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATG 358

Db        944 TGGCCCTCAACGAGCCCAACATG---GGCCGATCGGTGTGGGACGGACTGCTCACGAACC 888

Qy        359 TTCTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCA 418

Db        887 TGCTGCTGATCCAGGCATGGTTCCCCGACCACCACGAGTACGGCAGCATGAACCCGGTGG 828

Qy        419 GCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGG 478

Db        827 CGTGGTCGCTCTCCTGCGAGCTGTTCTTCTACGCCATGTTCCCGTTCCTCTTCGCCTTCT 768

Qy        479 TGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGA 538
```

```
Db        767 TCACCAAGGTCCGTACGGACCGGCTCTGGCGGTGGGCCGCCGCGGTGTCCGTGGCCGCCG 708

Qy        539 CGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAG------CGAGGAGGTGGCCCCCG 592

Db        707 TCTCCATCCCCCTGGTCGCACTGCTGCTGCCGGCCAGCCCGCCCCTGCCGTGGGACCCGG 648

Qy        593 GGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGT 652

Db        647 ACATGCCGCAGCTGCGGTGGTGGTTCATCTACATGTTCCCGCCGGTGCGGCTGCTGGAGT 588

Qy        653 TCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712

Db        587 TCGTGCTCGGGATGCTCATGGCCCAGATCGTGATCCGGGGACGCTGGAGGGGCCCGCGTC 528

Qy        713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772

Db        527 CCCTGGCCTGCGTCGCGCTGTTCTCAGCGGTGTTCGCGGTGACGTTCGCGGTGCCGAACC 468

Qy        773 TCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCG 832

Db        467 ACTACGACCCCGGCGCGTTGACCGTCCCGGTGATCGCGCTGCTGCTCGCCTCGGTGGCCG 408

Qy        833 ACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCC 892

Db        407 TCGGTGATGTGCGCGGCGTCCGCTCCTGGCTGGGGACCAGGACGATGGTGCTGCTGGGGG 348

Qy        893 AGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGA 952

Db        347 AACTCACCTTCGCCTTCTACCTCGTGCACTACCTGATCATCCAGTACGGGCACCGCTTCG 288

Qy        953 TGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCC 1012

Db        287 CCGGCGGGAAGCAGGGCTATTACCGGCAGTGGGACACACCGGCCGCCGTCGGGCTGACCC 228

Qy       1013 TGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGC 1072

Db        227 TGCTCGCCTTCACGCTGGCGCTGGGGCTGTCGGCGTTCCTGCACTTCTTCGTGGAGAAGC 168

Qy       1073 CCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG 1132

Db        167 CGGTCATGCGAACCCTGGGACGGC-CGCGGCGGTCCCCGGACGCCGGCTCGACACCCAGG 109

Qy       1133 CGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTAC 1192

Db        108 TCCGAACCCGCCCCGTCCGGCACTCCGTAGCCGACGCGGGACGACCGGTGCGGGGCGCGC 49

Qy       1193 CCGGGCC 1199

Db         48 CCTCGCC 42


RESULT 9
AR363942/c
LOCUS       AR363942               2749 bp    DNA     linear   PAT 03-SEP-2003
DEFINITION  Sequence 1 from patent US 5240849.
ACCESSION   AR363942
VERSION     AR363942.1  GI:34426061
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2749)
  AUTHORS   Arisawa,A., Kawamura,N., Kojima,I., Okumura,Y., Okamura,K., Tone,H.
            and Okamoto,R.
  TITLE     DNA coding for enzyme capable of acylating the 4'-position of
            macrolide antibiotic
  JOURNAL   Patent: US 5240849-A 1 31-AUG-1993;
            Sanraku Incorporated; Tokyo;
            JPX;
FEATURES             Location/Qualifiers
     source          1. .2749
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           28.5%;  Score 355.4;  DB 2;  Length 2749;
  Best Local Similarity   58.2%;  Pred. No. 2.4e-73;
  Matches  703;  Conservative    0;  Mismatches  491;  Indels   13;  Gaps    4;

Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61

Db       1244 TGCCCCTGCCGAAACATCTTCCCGCGCTCGGCGGGATGCGTTTCATCTCCGCTCTACTGG 1185

Qy         62 TATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCAC 121

Db       1184 TATTCACCTCCCATATATCGACACAGCCGTTCTTCAAGAACACCGAGATCAATTCCGCGC 1125

Qy        122 TGC---TGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTG 178

Db       1124 TGCAGTTCCCGCTGAACCGGCTGGGCCCGCTGACGGTCTCGTTCTTCTTCATGCTCAGCG 1065

Qy        179 GATTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCC 238

Db       1064 GTTTCGTCCTCACCTGGGCGGGTCTGCCCGACAAGTCCAAGGTGAACTTCTGGCGGCGGC 1005

Qy        239 GCATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTT 298

Db       1004 GCACGGTCCGCGCGTACTCGCTGCACCTGCCCGTGCTGCTGGTGACGCTGCTGATCGTGC 945

Qy        299 TCTCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATG 358

Db        944 TGGCCCTCAACGAGCCCAACATG---GGCCGATCGGTGTGGGACGGACTGCTCACGAACC 888

Qy        359 TTCTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCA 418

Db        887 TGCTGCTGATCCAGGCATGGTTCCCCGACCACCACGAGTACGGCAGCATGAACCCGGTGG 828

Qy        419 GCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGG 478

Db        827 CGTGGTCGCTCTCCTGCGAGCTGTTCTTCTACGCCATGTTCCCGTTCCTCTTCGCCTTCT 768

Qy        479 TGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGA 538

Db        767 TCACCAAGGTCCGTACGGACCGGCTCTGGCGGTGGGCCGCCGCGGTGTCCGTGGCCGCCG 708

Qy        539 CGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAG------CGAGGAGGTGGCCCCCG 592

Db        707 TCTCCATCCCCCTGGTCGCACTGCTGCTGCCGGCCAGCCCGCCCCTGCCGTGGGACCCGG 648

Qy        593 GGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGT 652

Db        647 ACATGCCGCAGCTGCGGTGGTGGTTCATCTACATGTTCCCGCCGGTGCGGCTGCTGGAGT 588

Qy        653 TCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712

Db        587 TCGTGCTCGGGATGCTCATGGCCCAGATCGTGATCCGGGGACGCTGGAGGGGCCCGCGTC 528

Qy        713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772

Db        527 CCCTGGCCTGCGTCGCGCTGTTCTCAGCGGTGTTCGCGGTGACGTTCGCGGTGCCGAACC 468

Qy        773 TCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCG 832

Db        467 ACTACGACCCCGGCGCGTTGACCGTCCCGGTGATCGCGCTGCTGCTCGCCTCGGTGGCCG 408

Qy        833 ACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCC 892

Db        407 TCGGTGATGTGCGCGGCGTCCGCTCCTGGCTGGGGACCAGGACGATGGTGCTGCTGGGGG 348

Qy        893 AGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGA 952

Db        347 AACTCACCTTCGCCTTCTACCTCGTGCACTACCTGATCATCCAGTACGGGCACCGCTTCG 288

Qy        953 TGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCC 1012
```

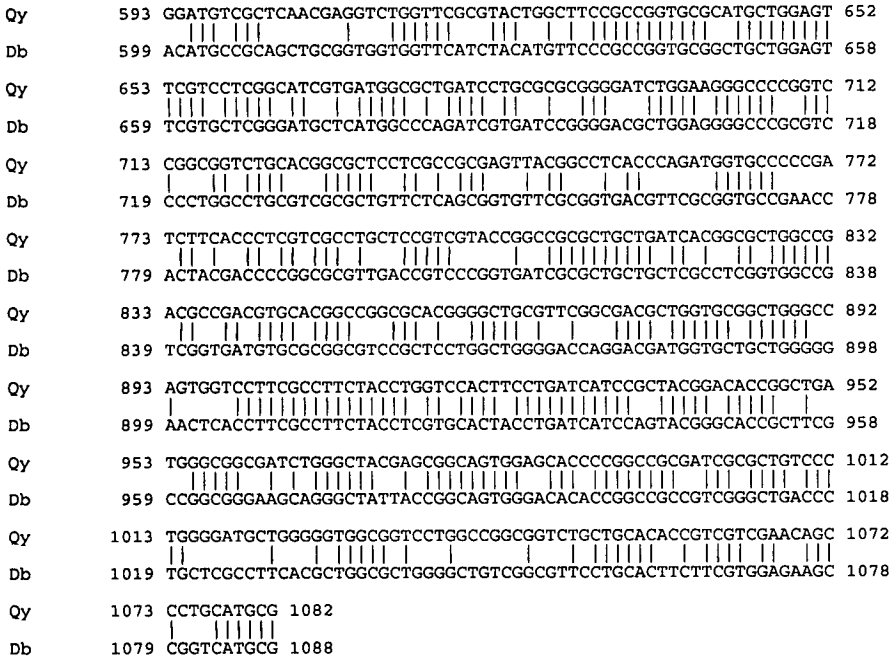
```
Db        287 CCGGCGGGAAGCAGGGCTATTACCGGCAGTGGGACACACCGGCCGCCGTCGGGCTGACCC 228

Qy       1013 TGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGC 1072

Db        227 TGCTCGCCTTCACGCTGGCGCTGGGGCTGTCGGCGTTCCTGCACTTCTTCGTGGAGAAGC 168

Qy       1073 CCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG 1132

Db        167 CGGTCATGCGAACCCTGGGACGGC-CGCGGCGGTCCCCGGACGCCGGCTCGACACCCAGG 109

Qy       1133 CGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTAC 1192

Db        108 TCCGAACCCGCCCCGTCCGGCACTCCGTAGCCGACGCGGGACGACCGGTGCGGGGCGCGC 49

Qy       1193 CCGGGCC 1199

Db         48 CCTCGCC 42


RESULT 10
STMACYB/c
LOCUS       STMACYB                 2749 bp    DNA     linear   BCT 07-FEB-1999
DEFINITION  Streptomyces thermotolerans carB, acyB1 or carE, and acyB2 genes
            for carbomycin resistance detarminant, 4'-mycarosyl isovaleryl-CoA
            transferase and regulatory protein, partial and complete cds.
ACCESSION   D31821
VERSION     D31821.1  GI:506870
KEYWORDS    4'-mycarosyl isovaleryl-CoA transferase; carbomycin resistance
            determinant; regulatory protein.
SOURCE      Streptomyces thermotolerans
  ORGANISM  Streptomyces thermotolerans
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (sites)
  AUTHORS   Epp,J.K., Burgett,S.G. and Schoner,B.E.
  TITLE     Cloning and nucleotide sequence of a carbomycin-resistance gene
            from Streptomyces thermotolerans
  JOURNAL   Gene 53 (1), 73-83 (1987)
   PUBMED   3036668
REFERENCE   2  (sites)
  AUTHORS   Epp,J.K., Huber,M.L., Turner,J.R., Goodson,T. and Schoner,B.E.
  TITLE     Production of a hybrid macrolide antibiotic in Streptomyces
            ambofaciens and Streptomyces lividans by introduction of a cloned
            carbomycin biosynthetic gene from Streptomyces thermotolerans
  JOURNAL   Gene 85 (2), 293-301 (1989)
   PUBMED   2628170
REFERENCE   3  (bases 1 to 2749)
  AUTHORS   Arisawa,A., Kawamura,N., Tsunekawa,H., Okamura,K., Tone,H. and
            Okamoto,R.
  TITLE     Cloning and nucleotide sequences of two genes involved in the
            4''-O-acylation of macrolide antibiotics from Streptomyces
            thermotolerans
  JOURNAL   Biosci. Biotechnol. Biochem. 57 (12), 2020-2025 (1993)
   PUBMED   7764361
REFERENCE   4  (bases 1 to 2749)
  AUTHORS   Arisawa,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1994) Akira Arisawa, Mercian Corporation Central
            Laboratories; 4-9-1 Johnan, Fujisawa-shi, Kanagawa 251, Japan
            (E-mail:mercian2@city.fujisawa.kanagawa.jp, Tel:0466-35-1511,
            Fax:0466-35-1530)
FEATURES             Location/Qualifiers
     source          1. .2749
                     /organism="Streptomyces thermotolerans"
                     /mol_type="genomic DNA"
                     /strain="TH475"
                     /db_xref="taxon:80858"
                     /clone="pMAB60"
     gene            1. .34
                     /gene="carB"
     CDS             <1. .34
                     /gene="carB"
                     /citation=[1]
                     /codon_start=2
                     /transl_table=11
                     /product="carbomycin resistance determinant"
                     /protein_id="BAA06606.1"
                     /db_xref="GI:506871"
                     /translation="ITLFEELHGR"
     unsure          39. .40
                     /citation=[2]
                     /replace="caggcctg"
     unsure          43. .44
                     /citation=[2]
                     /replace="gggcgcgccc"
     unsure          56
                     /citation=[2]
                     /replace="g"
     gene            complement(79. .1245)
                     /gene="acyB1 or carE"
     CDS             complement(79. .1245)
                     /gene="acyB1 or carE"
                     /function="4'-O-acylation of macrolide antibiotics"
                     /citation=[2]
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="4'-mycarosyl isovaleryl-CoA transferase"
                     /protein_id="BAA06607.1"
                     /db_xref="GI:506872"
                     /translation="MPLPKHLPALGGMRFISALLVFTSHISTQPFFKNTEINSALQFP
                     LNRLGPLTVSFFFMLSGFVLTWAGLPDKSKVNFWRRRTVRAYSLHLPVLLVTLLIVLA
                     LNEPNMGRSVWDGLLTNLLLIQAWFPDHHEYGSMNPVAWSLSCELFFYAMFPFLFAFF
                     TKVRTDRLWRWAAAVSVAAVSIPLVALLLPASPPLPWDPDMPQLRWWFIYMFPPVRLL
                     EFVLGMLMAQIVIRGRWRGPRPLACVALFSAVFAVTFAVPNHYDPGALTVPVIALLLA
                     SVAVGDVRGVRSWLGTRTMVLLGELTFAFYLVHYLIIQYGHRFAGGKQGYYRQWDTPA
                     AVGLTLLAFTLALGLSAFLHFFVEKPVMRTLGRPRRSPDAGSTPRSEPAPSGTP"
     RBS             complement(1251. .1255)
     -10_signal      complement(1281. .1287)
     unsure          1304
                     /citation=[2]
                     /replace=""
     -35_signal      complement(1306. .1311)
     -35_signal      1475. .1480
     -10_signal      1493. .1498
     RBS             1559. .1562
     gene            1569. .2732
                     /gene="acyB2"
     CDS             1569. .2732
                     /gene="acyB2"
                     /function="activation of acyB1 expression"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="regulatory protein"
                     /protein_id="BAA32409.1"
                     /db_xref="GI:3434935"
                     /translation="MHSIPCGSKPSASMWDTGVHDDFDTHISETCSELFSSLRRADQR
                     KRGEQYVRGLLTAQGRKTARNLAAFVGEGAADQNLHHFVAGSTWDWRSVRAALARYAD
                     QTVRGDAWVIRPMVVYKAGGRSVGVGRRFVPDLGRVVSCQQSYGLWLASDAMSAPVNW
                     HLTLGGGPGDRHDRQLSAYGEEEKLVDLVAELTRSNRVLARPVVMDARIATLPRLVRA
                     LSAADQSFLLRVSGDLPLALAGSRGQLDRRAQVWPAQHLMEQLKRLRRPVEWQGSISF
                     VAPCNVVLTDQLPQRTLLLFGVWRANRRRPADLWLTDLTSWNRGALLRLAMLTCRVDA
                     DFARVSLGVGIRDFEGRSFQGWHRHVTLASIAHALRLSCTDTARTPTAPALSR"
ORIGIN

  Query Match             28.5%;  Score 355.4;  DB 15;  Length 2749;
  Best Local Similarity   58.2%;  Pred. No. 2.4e-73;
  Matches  703;  Conservative    0;  Mismatches  491;  Indels   13;  Gaps    4;

Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61

Db       1244 TGCCCCTGCCGAAACATCTTCCCGCGCTCGGCGGGATGCGTTTCATCTCCGCTCTACTGG 1185

Qy         62 TATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCAC 121
```





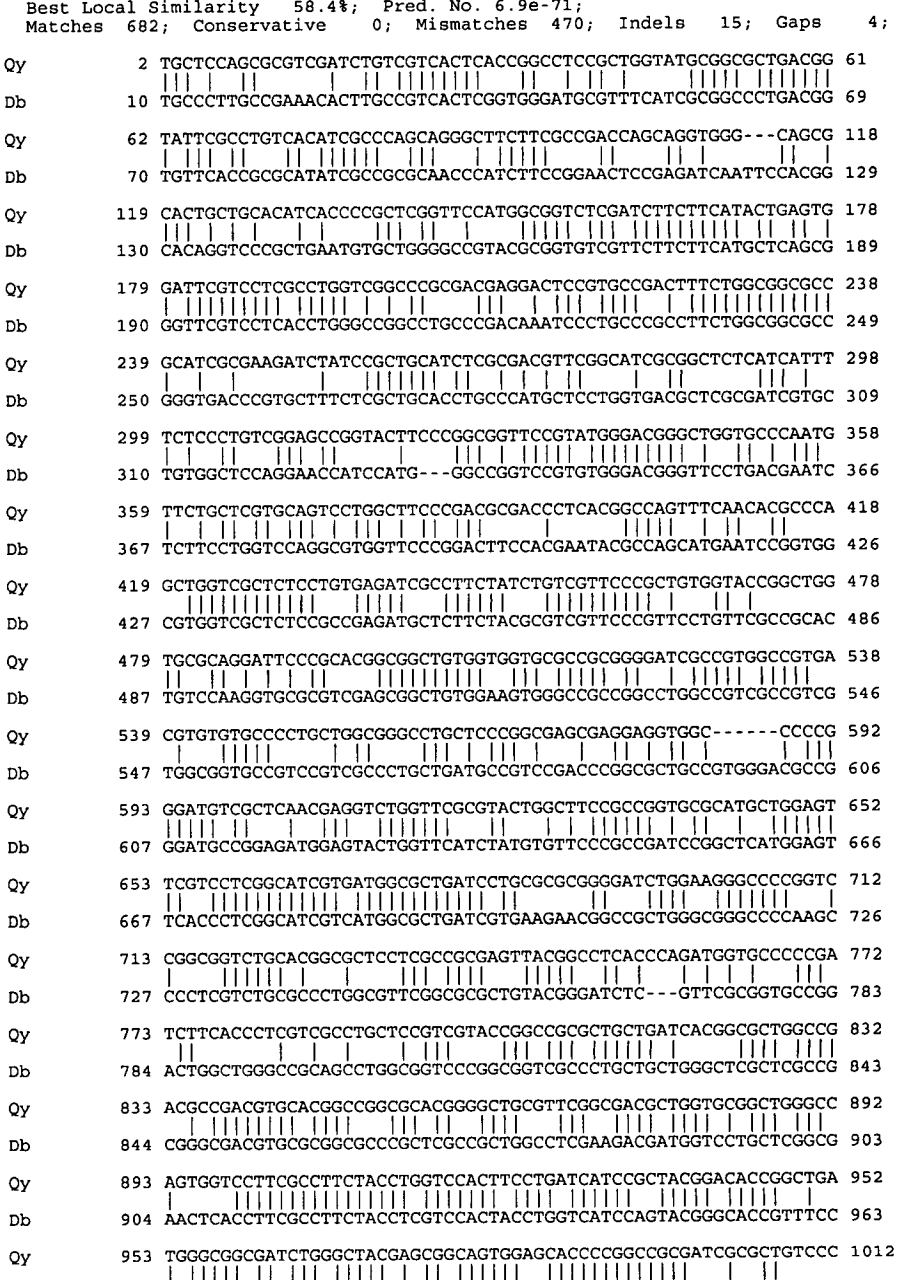
RESULT 11
AR363943
LOCUS AR363943 1167 bp DNA linear PAT 03-SEP-2003
DEFINITION Sequence 2 from patent US 5240849.
ACCESSION AR363943
VERSION AR363943.1 GI:34426062
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1167)
AUTHORS Arisawa,A., Kawamura,N., Kojima,I., Okumura,Y., Okamura,K., Tone,H. and Okamoto,R.
TITLE DNA coding for enzyme capable of acylating the 4'-position of macrolide antibiotic
JOURNAL Patent: US 5240849-A 2 31-AUG-1993;
Sanraku Incorporated; Tokyo;
JPX;
FEATURES Location/Qualifiers
source 1. .1167
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Query Match 28.2%; Score 351.2; DB 2; Length 1167;
Best Local Similarity 59.6%; Pred. No. 2.3e-72;
Matches 650; Conservative 0; Mismatches 428; Indels 12; Gaps 3;



RESULT 12
E28865
LOCUS E28865 1646 bp DNA linear PAT 18-JUN-2001
DEFINITION DNA containing macrolide antibiotic acyltransferase gene.
ACCESSION E28865
VERSION E28865.1 GI:13020905
KEYWORDS JP 1999285384-A/1.
SOURCE Streptomyces kitasatoensis
ORGANISM Streptomyces kitasatoensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 1646)
AUTHORS Kaoru,K. and Tsukasa,Y.
TITLE DNA containing macrolide antibiotic acyltransferase gene
JOURNAL Patent: JP 1999285384-A 1 19-OCT-1999;
ASAHI CHEM IND CO LTD
COMMENT OS Streptomyces kitasatoensis
PN JP 1999285384-A/1
PD 19-OCT-1999
PF 02-APR-1998 JP 1998089651
PR
PI KAORU KOGA,TSUKASA YAMAGUCHI
PC C12N15/09,C12N9/80//C12N1/21,(C12N15/09,C12R1:465),(C12N9/80,
PC C12R1:465),
PC (C12N1/21,C12R1:465),C12N15/00,(C12N15/00,C12R1:465) CC
Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT CDS 9. .1178.
FEATURES Location/Qualifiers
source 1. .1646
/organism="Streptomyces kitasatoensis"
/mol_type="genomic DNA"
/db_xref="taxon:97405"
ORIGIN

Query Match 27.7%; Score 345; DB 2; Length 1646;
Best Local Similarity 58.4%; Pred. No. 6.9e-71;
Matches 682; Conservative 0; Mismatches 470; Indels 15; Gaps 4;



```
Qy       1013 TGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGC 1072
Db       1024 TGCTCGCCTTCGCGCTGAGCCTCGCCCTCGCCGCCTTCCTCCACTTCTTCGTGGAGAAGC 1083

Qy       1073 CCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG 1132
Db       1084 CCGTCATGCGGACCATCGGCCGGCCGCGCCGCCCGGCCCCGCCCGCCCCGGACCCGGCCC 1143

Qy       1133 CGGCTCCCCGGAACTCACCCGCGGCCG 1159
Db       1144 CGCGGTCCGGCCCGCCGGCCGCCGCCG 1170
```

RESULT 13
STMMPT
LOCUS STMMPT 1592 bp DNA linear BCT 28-MAR-2000
DEFINITION Streptomyces mycarofaciens mpt gene for midecamycin 4''-O-propionyl transferase, complete cds.
ACCESSION D63662
VERSION D63662.1 GI:927227
KEYWORDS midecamycin 4''-O-propionyl transferase; mpt.
SOURCE Streptomyces mycarofaciens
ORGANISM Streptomyces mycarofaciens
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (sites)
AUTHORS Xulun,Z. and Yiguang,W.
TITLE Studies on midecamycin 4'' -O-propionyltransferase gene structure
JOURNAL Acta Microbiol. Sin. 36, 417-422 (1996)
REFERENCE 2 (bases 1 to 1592)
AUTHORS Xulun,Z. and Yiguang,W.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1592)
AUTHORS Xulun,Z.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-1995) Zhang Xulun, Chinese Academy of Medical Sciences and Peking Union Medical; Tianta, Beijing 100050, China (Tel:0086-010-3037740, Fax:0086-10-3017302)
FEATURES Location/Qualifiers
source 1. .1592
/organism="Streptomyces mycarofaciens"
/mol_type="genomic DNA"
/strain="1748"
/db_xref="taxon:1949"
/clone="pCN1OF5"
-35_signal 283. .288
-10_signal 307. .313
RBS 342. .346
gene 353. .1519
/gene="mpt"
CDS 353. .1519
/gene="mpt"
/experiment="experimental evidence, no additional details recorded"
/note="similar to CarE"
/codon_start=1
/transl_table=11
/product="midecamycin 4''-O-propionyl transferase"
/protein_id="BAA09815.1"
/db_xref="GI:927228"
/translation="MPLPKHLPSLGGMRAIAALVVFWSHIASQTFFRKAEINSTAQVP
LDVLGPVAVFFFMLSGFVLTWAGMPDPSKPAFWRRRWVRVYSLHLPVLLLTLAIVLWL
KEPNMGGSVWDGFLSNLLLVQSWCPDYHQYGSMNPVAWSLSCEMLFYAAFPFLFAFFS
KMRAERLWSWVLGISVVAAAVPALALLLPSAPTLPWDPNMPELQYWFIYMLRAVRLLE
FALGGVLMAQIVRRGRWIGPTPGVCALLFAGAFAVSFALPSYLARDGPTVPLIALLLG
SLAAGDIRGTRSWLGTRTMVLLGELTFAFYVIHYLVIQYGHRFLGGELSYYRQWDTPA
AIGLTVLALGLSVGLAALLHFFVEKPVVRALGRSGKASRASKAPQPEPPAPLLS"
terminator 1543. .1572
ORIGIN

Query Match 25.5%; Score 317; DB 15; Length 1592;
Best Local Similarity 55.5%; Pred. No. 3e-64;
Matches 677; Conservative 0; Mismatches 530; Indels 12; Gaps 3;

```
Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61
Db        354 TGCCCTTGCCGAAACACCTGCCGTCGCTCGGCGGCATGCGGGCCATCGCCGCACTGGTGG 413

Qy         62 TATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCAC 121
Db        414 TGTTCTGGTCTCATATCGCTTCCCAGACGTTTTTCCGCAAGGCCGAGATAAACTCCACCG 473

Qy        122 TGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGAT 181
Db        474 CACAGGTCCCGCTGGACGTCCTGGGACCGGTGGCGGTCTTCTTCTTCATGCTCAGCGGAT 533

Qy        182 TCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCA 241
Db        534 TCGTCCTCACCTGGGCGGGCATGCCCGACCCGTCCAAGCCTGCCTTCTGGCGCCGCCGTT 593

Qy        242 TCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCT 301
Db        594 GGGTTCGGGTCTACTCGCTGCACCTGCCGGTTCTGCTGCTCACGCTGGCGATCGTGCTGT 653

Qy        302 CCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTC 361
Db        654 GGCTGAAGGAACCCAATATG---GGCGGGTCGGTGTGGGACGGCTTCCTCAGCAACCTGC 710

Qy        362 TGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCT 421
Db        711 TGCTCGTCCAGTCGTGGTGCCCCGACTACCACCAGTACGGCAGCATGAACCCGGTGGCGT 770

Qy        422 GGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGC 481
Db        771 GGTCCCTCTCCTGCGAGATGCTGTTCTACGCCGCCTTCCCGTTCCTGTTCGCCTTCTTCT 830

Qy        482 GCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGT 541
Db        831 CCAAGATGCGTGCCGAGCGGCTGTGGTCCTGGGTCCTGGGCATCTCCGTCGTCGCCGCGG 890

Qy        542 GTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGT------GGCCCCCGGGA 595
Db        891 CCGTGCCCGCCCTGGCCCTGCTGCTCCCCTCGGCCCCCACGCTGCCCTGGGACCCGAACA 950

Qy        596 TGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCG 655
Db        951 TGCCGGAGCTCCAATACTGGTTCATCTACATGCTTCGCGCGGTGCGGCTGCTGGAATTCG 1010

Qy        656 ---TCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712
Db       1011 CGCTCGGCGGCGTCCTGATGGCGCAGATCGTCAGGCGCGGCCGCTGGATCGGCCCGACCC 1070

Qy        713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772
Db       1071 CTGGGGTGTGCGCGCTGCTGTTCGCCGGCGCGTTCGCCGTGTCCTTCGCCCTGCCGTCCT 1130

Qy        773 TCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCG 832
Db       1131 ATCTGGCTCGCGATGGCCCGACGGTCCCGCTGATCGCGCTGCTGCTGGGCTCCCTGGCAG 1190

Qy        833 ACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCC 892
Db       1191 CTGGCGACATACGCGGTACCCGGTCGTGGCTGGGCACCCGGACGATGGTGCTGCTGGGTG 1250

Qy        893 AGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGA 952
Db       1251 AACTCACCTTCGCCTTCTACGTCATCCACTACCTCGTCATCCAGTACGGGCACCGCTTCC 1310

Qy        953 TGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCC 1012
Db       1311 TCGGCGGTGAGCTGAGCTACTACCGACAGTGGGACACCCCGGCCGCGATCGGCCTCACCG 1370

Qy       1013 TGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGC 1072
Db       1371 TTCTCGCCCTCGGGCTCAGCGTGGGCCTGGCCGCGCTCCTCCACTTCTTCGTGGAGAAGC 1430
```

```
Qy       1073 CCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG 1132
Db       1431 CGGTCGTCCGGGCCCTCGGCCGCTCCGGCAAGGCGTCCCGCGCGTCCAAGGCCCCGCAGC 1490

Qy       1133 CGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTAC 1192
Db       1491 CCGAGCCGCCGGCGCCCCTGCTGTCCTGAGCGGGTCCGGCGGCACAACAGTGTGCGGGGC 1550

Qy       1193 CCGGGCCCGCGCACACCCC 1211
Db       1551 GCGCCACATGGCGCGCCCC 1569
```

RESULT 14
AY735112
LOCUS AY735112 48200 bp DNA linear BCT 22-SEP-2004
DEFINITION Streptomyces hygroscopicus mannopeptimycin biosynthetic gene cluster, complete sequence.
ACCESSION AY735112
VERSION AY735112.1 GI:52222852
KEYWORDS .
SOURCE Streptomyces hygroscopicus
ORGANISM Streptomyces hygroscopicus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 48200)
AUTHORS Magarvey,N.A., Haltli,B., He,M., Greenstein,M. and Hucul,J.A.
TITLE The molecular logic of mannopeptimycin assembly
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 48200)
AUTHORS Magarvey,N.A., Haltli,B., Greenstein,M. and Hucul,J.A.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-2004) Natural Products Discovery, Wyeth Research, 401 North Middletown Rd, Pearl River, NY 10965, USA

```
FEATURES             Location/Qualifiers
     source          1. .48200
                     /organism="Streptomyces hygroscopicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1912"
     gene            77. .1048
                     /gene="mppC"
     CDS             77. .1048
                     /gene="mppC"
                     /codon_start=1
                     /transl_table=11
                     /product="acetyltransferase MppC"
                     /protein_id="AAU34194.1"
                     /db_xref="GI:52222853"
                     /translation="MSATPRPRPVLRPFRPGDGRSLLAAWCRSAPDDPITAARFRTLI
                     LLDPNFDPEGLRVADLDGQVVGAVYAVRRRTPLAGTDLEPDVGWILFFFVDPPHRRTG
                     LGRRLLTDALDWLRGHGRTRVDFAPYAPHYVLPGLDRAAYPEAARLLASLGFRPRYEA
                     AAMDRGLVGYRMPDEVRRHEAALTARGHRFGTPSDDDLVDLLGLAEEFTPDWARAIRQ
                     CLTGGAPLDRIVSARAPDGRMAGWAMHGAYDGTAERFGPFGVRKELRGAGLGKVLLHL
                     TLERMRALGVHGAWFLWTGEQSPAGHLYRASGFTTTRRFTVLRWEAG"
     gene            1045. .2460
                     /gene="mppD"
     CDS             1045. .2460
                     /gene="mppD"
                     /codon_start=1
                     /transl_table=11
                     /product="sugar binding protein MppD"
                     /protein_id="AAU34195.1"
                     /db_xref="GI:52222854"
                     /translation="MRRRTFTAGAAAGAALLAGAGCDAPGGAGHGDGEHGDGDGGDGR
                     GSGGRRGAPVTLTVLTHYASEPLASALQTVVDAWNATHRRITVRTAAVKFPDLLTTYM
                     VRQAAGQGADIIHPYCLWTGQLVRAGVLRPVPPTATRQIRRDFTPAAVAASSVHGTLY
                     GYPTEVQTYALYYNKRLLRQAGIDGPPGTWQELEDAAYRTARRDRHGNMLVQGFGLSR
                     ADDASVVGQTLALLAARGGTFLTSDGRRTAIGSAAGRDVLDLERRLIDRGAADSGISL
                     LRAFPSGQVAMAINAGWWTASLRGAMGADYREVGVAPVPGPAPDDRGTLATGFLLGVN
                     AKSRYPGEAWEFLHWLNGVRAPAARPGRSAGGGVPVSRMSALQVSVGSMTGRADDMRA
                     LLGGDGERDADGRGGGDRNLGPFLDALRYAVPEPNGPRAQQAKSLLRKNIEDVWTGRA
                     SVDAALRTAGRQIDQELSRPY"
     gene            2495. .3406
                     /gene="mppE"
     CDS             2495. .3406
                     /gene="mppE"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter MppE"
                     /protein_id="AAU34196.1"
                     /db_xref="GI:52222855"
                     /translation="MASAGGGPVRAARRRQTAVAYLFLTPALLFFAVFLALPLLFAVL
                     LAQSRWAGFDLADIEPVGMANFTDLFARGSTFLTPVLTNTLLYAVGTVAIALIGALTL
                     ATCIDNLRFQGLWRTLYFLPIVTTVVAVGNVWKYMYAPGGLINGVLNGLGLHSVAFLQ
                     DPGTALPSVVVVQAWASMGTAILILTAGLKSIPEAYYEAAELDGAGAGTVFRRITLPL
                     LRPSLLFVCITQFITGLQSFALINVMTDDGGPGDATNVAALEMYQQAFRYGDWGIASA
                     AAFVLFLVIVAITVGQLWLFRRKGGES"
     gene            3403. .4293
                     /gene="mppF"
     CDS             3403. .4293
                     /gene="mppF"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter MppF"
                     /protein_id="AAU34197.1"
                     /db_xref="GI:52222856"
                     /translation="MSRSARRRPGRRRPWGSYAVVVAGAALTLVPFLDMLLTSFKGPG
                     EYGKLPYRFLPQAFDLSNYRAAMEQLDLPLLFRNSVIATAVITGSILVTSALAGYALA
                     KLRFPGREVIFRLVLSTMMFPPFLFFIPHFLILVHWPGAGGNDLLGRGGAGLTVSLAA
                     LVMPFLVSGFGIFLMRQFMVSIPDELLEAARIDGAGEFALWWRIVLPQTKPVAVTLAL
                     LTFVNAWNEYIWALLISTANPRLMTLPVGIQMLQSYLDPDRMVPVMMAGLVLSILPVL
                     LLFLLLQKHYLRGVMLSGLK"
     gene            4359. .5636
                     /gene="mppG"
     CDS             4359. .5636
                     /gene="mppG"
                     /codon_start=1
                     /transl_table=11
                     /product="polyprenyl mannose synthase MppG"
                     /protein_id="AAU34198.1"
                     /db_xref="GI:52222857"
                     /translation="MSSGFSWAVVATVVRVSDPSGGTMASDSSSPTPMPAVSLIVPTF
                     NEAANIDELLDGVCAAIPAGLEVEVLFVDDSTDDTPEVIEKAAARCPMPVSVLHREVP
                     EGGLGGAVVAGIARTSAPWIMVMDADLQHPPELLPQLIEAGERAAAELVVASRYAEGG
                     SRGGLAGGYRVAVSGASTALTKSLFPRLLRGVSDPMSGCFAIRREAVDRAVQEGETRQ
                     EGGLRPLGYKILLELAVRCRPRGVVEVPYEFGERFAGESKSTVREGLRFLRHLAELRT
                     SDKRARMVAFGLIGVSGFVPNLLALWALTGATTLHYAVAEVLANQLGVLWNFALLDFL
                     VYRSGKPGRGAGRLLGFAALSNADLLARIPLMMLFVEQAGMGPVPATVISLVVVFALR
                     FLLVDTLIYRRKGAAAKRAADAAVTGGQGERAA"
     gene            5822. .7234
                     /gene="mppH"
     CDS             5822. .7234
                     /gene="mppH"
                     /codon_start=1
                     /transl_table=11
                     /product="polyprenyl phospho-mannosyltransferase MppH"
                     /protein_id="AAU34199.1"
                     /db_xref="GI:52222858"
                     /translation="MTVVLLALSDRYGYNVDELYFRLLGEHGWAWGYTDQPPLVPALV
                     HATAQVLGDSVWAIRVPAALCAGAVVLLGALITAELGGTRRAQTLSALGLGSSFLVLS
                     VGHIMVTTTLDMLAWAAVLLFVLRALLRSEGKWWLWAGVVLGLALYAKYIVALLPVAL
                     LAGLALVGPRKVFRDRWLYAGIALALAIGSPNLIYQATHDFPQLQMADALGATDGPMN
                     RVIFVPSLVILLGPVLTVVWVAGLVKLLRDPAWRPVRALAPAFVVGVALTLYGGGRPD
                     YVGGFLIGLFAAGAVAADRWMGRRTSRRVLLCAGLAASAVLQVLMALPVLPQSSPFVP
                     LNNISLESVGWPRLAEQVRTAYEALPRQQRERAVVLADNLGEIGALDRYGHGLPAVFS
                     GHNELHKWGPPPERADVVVAVGVPRSRLAAGFTSCTVVGRVDNGVGVENAEQGRPITV
                     CHGRKASWARLWPSYHYLSG"
     gene            7293. .8822
                     /gene="mppI"
     CDS             7293. .8822
                     /gene="mppI"
                     /codon_start=1
                     /transl_table=11
                     /product="polyprenyl phospho-mannosyltransferase MppI"
                     /protein_id="AAU34200.1"
                     /db_xref="GI:52222859"
```

```
                     /translation="MTTSLDRDSRAAAAGPGVFRPAPMAWRPVAVVVAALAVLLFAFA
                     GEYGYHADELYFRLLGVHGFAWGYVDQPPLLPLAVRTSMEIFGDSMWAIRVPAVLCAA
                     AVTALGAMIAAELGGSRRAQTLTAFGVATSTMVLSFGHWILTTSFDTVAWAAVLLFVM
                     RVLLRGESKWWLWAGVVVGVALYAKYIVLLLPVALLVGLALVGPRKVFRDGKLYAGTA
                     LALVIGSPNLIYQATHDFPQLQMAEGLAGTDGEANRAMFATNLILLFGPALFVLCMIG
                     LVKLFRVPEWKPVRTLAVGYLAATAASYLIEGGRPDYTGGLLIALLAAGCVTADRWAG
                     ARKLRLSVLAVSLTLSTAVQMLLSLPVIPKSSLRDFQIASMALETVGWPRLVQQTEAA
                     YRALPAADRDRAIVLTENFGEAGALDHYGHGLPKVYSGHNELYHWGPPPQRAEVVVAV
                     GIDRNRLSADFTSCKVVDHIDNRLGIDNPEQGVPITVCHGPKKPWSALWPTYRHYNAY
                     L"
     gene            9012. .10025
                     /gene="mppJ"
     CDS             9012. .10025
                     /gene="mppJ"
                     /codon_start=1
                     /transl_table=11
                     /product="methyltransferase MppJ"
                     /protein_id="AAU34201.1"
                     /db_xref="GI:52222860"
                     /translation="MSTEVSEAQARRAVADIFNSTLASSAIGAAWELGALDELRENGK
                     LDVSDFAVRHDLHEPAVVGMFTALASVGIVRREGATVVVGPYFDEANHHRSLFHWLNQ
                     GSGELFRRMPQVLPNENRTGKFYQRDAGAISYACREISERYFDPAFWAAVDGLGYTPT
                     TVADLGSGSGERLIQIARRFPGVRGLGVDIADGAIAMAEKEVAAKGFGDQISFVRGDA
                     RTIDQVSARGEFAEVDLLTCFMMGHDFWPRENCVQTLRKLRAAFPNVRRFLLGDATRT
                     VGIPDRELPVFTLGFEFGHDMMGVYLPTLDEWDGVFEEGGWRCVKKHAIDSLSVSVVF
                     ELE"
     gene            10069. .18312
                     /gene="mppA"
     CDS             10069. .18312
                     /gene="mppA"
                     /codon_start=1
                     /transl_table=11
                     /product="mannopeptimycin peptide synthetase MppA"
                     /protein_id="AAU34202.1"
                     /db_xref="GI:52222861"
                     /translation="MGEWRDRRLDELFAEQAARTPERTAVVFEGRAVSYRELDARAER
                     LAAVLAGRGAGPERFIALLLPRSAELIVAILAVLKSGAGYIPIDPEYPADRIAYILGD
                     ARPVATITTAEVRDGLPDPDTGSGTDWLILDESGYEQEPAGARPQPAPAAPRSAENPA
                     YVIYTSGSTGRPKGVVIPHSNVGRLLSSTAHWYGFDEQDVWPLFHSFAFDVSVWEIWG
                     ALLHGGKLVVVPHAVTRAPADFLRLLVEERVTVLNQTPSAFYQLMAADRENPALGAQL
                     ALRYVVFAGEALDLGKLADWYERHDDRAPTLVNMYGITETTVHSSFLALDKEGAAGAT
                     GSAVGVALPDLTFHVLDEDLRPVPVGAEGELYVAGPGLARNYAGRPGLTAERFVACPF
                     GPPGARMYRSGDLVRPLPDGGLEYLRRSDDQVKIRGFRIELGEISHALAQDPSVDQAT
                     VVVRDEASGERRLVAYVVPAGSARPTPSRLRAALATRLPGYMVPTAFHVMPAFPLTAN
                     GKLDRRALPAPTRQDSVDADYAAPEGATEEALAAIWREVLGVEQIGADDDFFELGGDS
                     LSVVRALSRMRTGLGLRLTAAEFFATPTVRALAARRERGTIGAPEQIPAAPRTGALPL
                     SFTQQRFWLFHELDPGEVEYNVHSALRLRGTLDLPALRTALGGLIARHEPLRTTVVSD
                     DGRPTAVIAPPEGFPVPLTVEDLSALTGDDQEAAQRRLLAEEVARPFDLAAGPVLRVL
                     VIRRGERDHALVIGVHHLATDGWSMGLLTDELSARYDAARRGVPAALEPLPVHYSDYA
                     AWQRATVDDGRLVPQIDYWRDRLADVAPLQLPTDRPRPALKTSAGAAHRFTLDRRLVA
                     ALKELSAAHGATLFMTLTAALQVLLARYSGQQDIALGTAVSGRDHPQVERLVGAFINT
                     VVLRSDVRGELPFHEFLGEVRETVLGAFAHQDLPFDRLVDALGAERDPSRTPLVQAML
                     LLQNAPAGAEEFAGLRTETVALPRPAAIFDLTVDCTERAGALEVMVEYNTDLFDATTI
                     ERLSGHLRVLLDAVCAAPRRQVRDLPLLPAAERDTLLTGWNDTAAALPTTLGVHRQFA
                     ERARTTPDALAVTHCGQTLTYAQLDARANQLAHYLGALGVGRGTPVVLNLARKPQLIV
                     AMLAVLKAGGAYVPTALDTPAARLGHLLEETGTPVLLTTARQAGALPPTEASVIDLDA
                     AGPDIARHPEHDPQVATRPEDLAYIVYTSGSTGRPKGVAVPHSALTDYCAWHNDALDV
                     GPEDRGSSVVGLAFDVAVGEVWPYLCAGARVDQPDQETLDDPTALVEWFAENGTTVAY
                     LPTPRIESLLDVAAITTTRLRTVLVIGDSLRRRPQPGLPFTLLNAYGPAEATVAATQA
                     VVEPLGPDAPAGLPSIGAPLYNTAAYVLDDRLCPVPVGVPGELYLAGAGLAQGYQGRP
                     DLTAERFVGCPFGPPGTRMYRTGDIVRWLPDGTLDFLGRIDNQVKLRGYRIELGEIES
                     VLARREELSQVFVTVREPSPGRRSLVAYLVADRGTAPDPEELAGYIASVLPEYMVPSS
                     FVLLDALPLTANGKIDRRALPEPEPAGGEGAAYVAPGNEVEETLAAIWAEVLGVERVG
                     VQDNFFALGGDSISGLQTAVRARRAGLRLASKDLFQRQTIAALSPVVTVERTTADADP
                     APSDRPTAPFALSGLDRVGVERLTADGGPAEDAYPLTPMQSGLLFHTLMHAERGMYIE
                     QFHFALHSIREPELLATAWQRVVDRTPVLRTSLAWDGLAEPLQVVRTGVRIPVAQLDW
                     TALDEAGQRQALERYLTEDRTRGLDLHTAPLARIAVARLGGDQVRLVWTFHHLLLDGW
                     SVVQVLSEVLGEYAALADGIPYTPQLRHTYAEFVGQLAGQDHTAAEKYWRAALTGRES
                     PTPLPYDRPRPDAHQAAPDAELKLRLPAAVTGRLGTAAKRAGVTMNTVVQGLWALLLA

Query Match          19.8%;  Score 246.6;  DB 15;  Length 48200;
Best Local Similarity  55.0%;  Pred. No. 1.9e-47;
Matches  611;  Conservative    0;  Mismatches  464;  Indels   36;  Gaps    5;

Qy        17 ATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACA 76
Db     33645 ACCTGAGATCCCTGACCGGTTTGCGGTTCGTGGCGATGCTGCCGGTCTTCCTCACCCATG 33704

Qy        77 TCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCC 136
Db     33705 CGGCGTTCGAGGGCGTCTTCAGCGACGCGAAGGTGAGCTGGGGCTTCCTCGACGCGATGG 33764

Qy       137 CGCT---CGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCT 193
Db     33765 GGAGCACCGGCTATATGGCCGTCTCGTTCTTCTTCGTGCTCAGCGGCTTTGTGATCACGT 33824

Qy       194 GGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCT 253
Db     33825 GGTCGTACCGGCCCACCGACACCGCGCGCAAGTTCTGGCGCCGGCGCTTCTTCCGGGTCT 33884

Qy       254 ATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGC 313
Db     33885 TCCCCAACCACGTCGTGACCTATGCGCTCGC------------CCTCGGGCTGATCGCTG 33932

Qy       314 CGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGT 373
Db     33933 CGGTGGGGCTGAGTGTCGGCGTACTGCCC---TCGGTCACCCAGCTCTTCCTCGTCCAGT 33989

Qy       374 CCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCT 433
Db     33990 CCTGGGTGCCCGACCCGGCGTTCACCGACACCGGCAACAGCGTGAGCTGGTCGCTCGCGG 34049

Qy       434 GTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCG 493
Db     34050 TGGATGTGGTGTTCTACGCGCTCTTCCCGGTGCTGCTCACGCTGGTGAACAAGATCAAGC 34109

Qy       494 CACGGCGGCTGTGGTGGTGCGCCG------CGGGGATCGCCGTGGCCGTGACGTGTGTGC 547
Db     34110 CGAATCGGCTCTGGTACTGGGTCGGTGGCTCCGTCATCGGTGTGGCCGTGGTACCGGCCA 34169

Qy       548 CCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACG 607
Db     34170 TCGCGCTCGCCGCGCTCCCGAGCACCCCCGAGATGCCGCTCGGCGGGGTGTCCGTCAGCC 34229

Qy       608 AGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCG 667
Db     34230 AGTACTGGTTCACCTACTTCTTCCCGCTCTTCCGGCTGCTGGAGTGTGTGCTCGGCATGC 34289

Qy       668 TGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGG 727
Db     34290 TGATGGCGCGGATCGTGCTGTCCGGCAAGTGGATACGCCTGCGGGTGCTGCCCGCCGCCG 34349

Qy       728 CGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCG 787
Db     34350 TCCTCGTGGTGATCGCGTACTACTTCGCCCAGCAGGTCCCGTACCTCTACCGGCTGAGTG 34409

Qy       788 CCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACG 847
Db     34410 CGGTGACGGTGCTGCCGGTCGCGCTGCTGACGGCGGCGGCCGCGGTGGCGGACTCCGAGG 34469

Qy       848 GCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCT 907
Db     34470 GCCGGGGCACCCTGTTCGGCAGCAAGGTCATGGTCTGGTTCGGCGAACTCTCCTTCGCCT 34529

Qy       908 TCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGG 967
Db     34530 TCTACCTGCTGCACAACCTCGTCCTGAAGTACGGCCATCTGCTGCTCGGCCACACCGAGG 34589

Qy       968 GCTACGAGCGGCAG------------TGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGG 1015
Db     34590 AGGAGGGCGAGCTGGTGGGCCACACCTGGGGCGTGCCCGAGGGAATCGCCCTGATCGCCG 34649

Qy      1016 GGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCT 1075
Db     34650 CCGCCTTCGCGGTGTCCCTGCTGCTGGCCTGGCTGCTGCACAACGGAGTGGAGAAGCAGG 34709
```

```
Qy       1076 GCATGCGCCTGTTCGGCAGCCGCAGGTCCGC 1106
              ||||||| |    | | |||| | | ||
Db      34710 CGATGCGCCGCTGGTCCCGACGCAAGCCGGC 34740


RESULT 15
AB088224
LOCUS       AB088224               210614 bp    DNA     linear   BCT 11-JUN-2003
DEFINITION  Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.
ACCESSION   AB088224
VERSION     AB088224.1  GI:30698345
KEYWORDS    .
SOURCE      Streptomyces rochei
  ORGANISM  Streptomyces rochei
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Kinashi,H., Fujii,S., Hatani,A., Kurokawa,T. and Shinkawa,H.
  TITLE     Physical mapping of the linear plasmid pSLA2-L and localization of
            the eryAI and actI homologs
  JOURNAL   Biosci. Biotech. Biochem. 62, 1892-1897 (1998)
REFERENCE   2
  AUTHORS   Suwa,M., Sugino,H., Sasaoka,A., Mori,E., Fujii,S., Shinkawa,H.,
            Nimi,O. and Kinashi,H.
  TITLE     Identification of two polyketide synthase gene clusters on the
            linear plasmid pSLA2-L in Streptomyces rochei
  JOURNAL   Gene 246 (1-2), 123-131 (2000)
   PUBMED   10767533
REFERENCE   3
  AUTHORS   Hiratsu,K., Mochizuki,S. and Kinashi,H.
  TITLE     Cloning and analysis of the replication origin and the telomeres of
            the large linear plasmid pSLA2-L in Streptomyces rochei
  JOURNAL   Mol. Gen. Genet. 263 (6), 1015-1021 (2000)
   PUBMED   10954087
REFERENCE   4
  AUTHORS   Mochizuki,S., Hiratsu,K., Suwa,M., Ishii,T., Sugino,F., Yamada,K.
            and Kinashi,H.
  TITLE     The large linear plasmid pSLA2-L of Streptomyces rochei has an
            unusually condensed gene organization for secondary metabolism
  JOURNAL   Mol. Microbiol. 48 (6), 1501-1510 (2003)
   PUBMED   12791134
REFERENCE   5  (bases 1 to 210614)
  AUTHORS   Mochizuki,S., Hiratsu,K. and Kinashi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2002) Haruyasu Kinashi, Hiroshima University,
            Department of Molecular Biotechnology, Graduate School of Advanced
            Sciences of Matter; 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima
            739-8530, Japan (E-mail:kinashi@hiroshima-u.ac.jp,
            Tel:81-824-24-7869, Fax:81-824-24-7869)
COMMENT     The nucleotide sequence has been determined by using restriction
            fragments and nested deletion fragments of the ordered cosmid
            library of pSLA2-L.
            pSLA2-L is 210,614 bp long with G+C content of 72.8 % and carries
            143 ORFs.
            Gene prediction was based on the unique codon usage in Streptomyces
            (Bibb et al., Gene 30:157-66 (1984)) using the FramePlot program of
            Ishikawa and Hotta (FEMS Microbiol Lett 174: 251-253 (1999)) as
            implemented at
            http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl. Where possible we
            chose an initiation codon (atg, gtg, ctg or ttg) which is preceded
            by an upstream ribosome binding site sequence (optimally 5-13 bp
            before the initiation codon). If this could not be identified we
            chose the most upstream initiation codon.
            Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes
            (three PKS gene clusters for lankacidin, lankamycin and an unknown
            type II polyketide, a carotenoid biosynthetic gene cluster, many
            regulatory genes and others). The range of each biosynthetic gene
            cluster has been deduced by comparison with similar gene clusters
            in most cases and may be revised in future.
FEATURES             Location/Qualifiers
     source          1. .210614
                     /organism="Streptomyces rochei"
                     /mol_type="genomic DNA"
                     /strain="7434AN4"
                     /db_xref="taxon:1928"
                     /plasmid="pSLA2-L"
                     /note="linear plasmid"
     LTR             1. .1992
                     /note="left terminal inverted repeat, TIR-L; shows 99.4 %
                     (1981/1992) sequence identity to TIR-R (complement
                     (208623. .210614))"
     CDS             683. .2188
                     /note="N-terminal sequence is almost identical (435/437)
                     with that of ORF143 at the right end of pSLA2-L until the
                     inner end of TIR-L
                     ORF1 (501 aa)
                     similar to AL590463 Streptomyces coelicolor putative
                     helicase, SCP1.136 (879 aa); homology is seen untill the
                     inner end of TIR-L"
                     /codon_start=1
                     /transl_table=11
                     /product="putative helicase"
                     /protein_id="BAC76459.1"
                     /db_xref="GI:30698346"
                     /translation="MSTTSRTDQREAAQQEAVDAVVRALELPVRLLAPERGLRTQVIM
                     ATGSGKTRVAARSAEKLRAGRVLVLVPSLDLLTQTEAAWREAGRTGPMIGVSSLRGED
                     VAFPNTTDVEELVDWVRPFDKVTVFATYASLGLGTLERAHRGGLPGWDLIVVDEAHRT
                     SGRLGKPWAVVHDNTRIPSLRRLYMTATPRLWQLDEDAEGAPGELVASMEDDPDGLFG
                     ARCFTLTLSEAIDRGICAPYQVVCVDITDTQLQAAQLLGVEGRSDEVRGARLAALQTA
                     LLKASSEENFRRTLVFHHMVKEAEAFAAGLPDVAKRLHAAGPGLYPRTIWANWLCGEH
                     KPGHRRRVLGEFTSGIATDGTVVEKGFLGSVKVLGEGVDTRECDSVYWADVRGSMPDL
                     VQAVGRALRIQPGQGKVASLVVPVLLDPGETADNMLTSRPYNGLALLFRQAPLLTGHG
                     EEGFRAARFPGFEVRPVLGVRGAQRGASAGPASRSITPSAMSSLSISMRAP"
     CDS             complement(2315. .3595)
                     /note="ORF2 (426 aa)
                     similar to AE004736-10 Pseudonomas aeruginosa hypothetical
                     protein (442 aa)"
                     /codon_start=1
                     /transl_table=11
                     /product="putative secreted protein"
                     /protein_id="BAC76460.1"
                     /db_xref="GI:30698347"
                     /translation="MRRKPRALLGIPLGLLVAVLAPAPAAHASADTYVGTWAAPPTAA
                     PAADSTVYEDQTLRQTVHLSVAGDSLRVRFTNEFGTSPLTIGEVHAARPAAQGPATAV
                     DPGTDRVVRFGGRPSATLAPGTQRWSDPVALPTTAGGDLVISLYLPQRTPGSTVHSAA
                     YQHNFVAAGDVTGAPDLTPVSTATSWHFLSGVSVDRRAGTADSSVVTLGDSITDGEHT
                     TLDADRRWPDLLAERLRRDGGLAGTGVVNAGIGGNRLLRDPDPEPGSAAESFAAYFGE
                     SALKRFDRDVLGQPGARAVTVLLGVNDLGQPGIAAPASDEVTAEELIAGYRQLIERAH
                     EHGLKIYGATITPFAGDTIGYFTPRREAVRQQVNDWIRTSGAFDTVLDFDAVLRDPAR
                     PDHLLPAYDGGDGLHPNDAGMAAMARAFPLDSLR"
     CDS             complement(3659. .4645)
                     /note="ORF3 (328 aa)
                     similar to Y00459-2 Streptomyces griseus regulatory
                     protein, StrR (350 aa)"
                     /codon_start=1
                     /transl_table=11
                     /product="StrR-like regulatory protein"
                     /protein_id="BAC76461.1"
                     /db_xref="GI:30698348"
                     /translation="MTAALAGNLDQETRSIPISSLVPGYSPRLCGEDAEHVARLADTD
                     EALPPILVERSTLRVIDGMHRVLAAKAKGHTTIEVRLFDGAAEEAFLLAVRSNMTHGL
                     PLSRQDRRAAAQRILAQWPHLSDRAVAGIAGIGAKTVAALRPLAAGPTPNPQARRGRD
                     GRIRPLDGTTGRRKAAELLAQRPQASVREVARHAGISPATASDVRRRLASGRSPVPER
                     NTPGARPAGGTGSRATPGGGAAPEPVSVVRPIRPPEDSPLVRLLRDPSLRHKESGRRL
                     LRLLQCGAVERTALLAMAQTVPPHCTDLVAELAREYADLWAEFAREVARTDG"
     CDS             complement(4922. .5812)
                     /note="ORF4 (296 aa), lankacidin biosynthesis protein
                     similar to AE004625-4 Pseudomonas aeruginosa
                     pyrroloquinoline quinone biosynthesis protein B (304 aa)"
                     /codon_start=1
                     /transl_table=11
                     /product="pyrroloquinoline quinone biosynthesis protein B"
                     /protein_id="BAC76462.1"
                     /db_xref="GI:30698349"
                     /translation="MILLGTAAGGGFPQWNCACALCARGRRGELPARSQECVAVSGDG
                     RDWWLLNASPDIRTQLLAAPALTPGPGPRDTPVRGVLLTDAEVDHALGLAVLRGATGL
```

```
                TVYAAPPVRGALSAELPVRGLLDRYAPWDWRDATAPGGFAVAGGLTVTAHPVGTKAPK
                YAHAPDPDAPWVCAYRIEDPATGGALVYAPCLATWPDGFDDLLASATCALLDGTFFSA
                GELGTATSSAGAGQSLMGHLPVAGPGGSLAALARHRGLRRIYTHLNNTNPLLDPSSAA
                HAAVREAGVEVLPDGSELVL"
     CDS        complement(5815. .6909)
                /note="ORF5 (364 aa), lankacidin biosynthesis protein
                similar to AJ277117-5 Gluconobacter oxydans putative pqqE
                protein (359 aa)"
                /codon_start=1
                /transl_table=11
                /product="pyrroloquinoline quinone biosynthesis protein E"
                /protein_id="BAC76463.1"
                /db_xref="GI:30698350"
                /translation="MADPAVGAPAGMLIELTHRCPLHCPYCSNPLELVRREAELTCEQ
                WTDILTQARELGVVQMHFSGGEPLARPDLPDLVGHARRLGAYVNLVTSGVGLTAERAH
                DLARRGVDHVQLSLQDADPAAGQAIAGARVHTAKLEAARAVTAAGLPLTVNIVLHRGN
                IDRTGRMVDLAVDLGADRIELANTQYYGWGLRNRAALMPTAAQLAAAREAVRHARTRY
                AGGPELVYVAADYYDDRPKPCMDGWGSTQLTVTPAGDVLPCPAAYAITTLPVENALRR
                PLSEIWYASRSFNAYRGTGWMREPCRTCPERHADHGGCRCQAFQLTGDAAATDPACGL
                SPHRSLVDAALAEVTDGPVPAFVPRGPVPA"
     CDS        complement(6902. .7174)
                /note="ORF6 (90 aa), lankacidin biosynthesis protein
                similar to AL603642-200 Sinorhizobium meliloti putative
                pyrroloquinoline quinone biosynthesis protein D (98 aa)"
                /codon_start=1
                /transl_table=11
                /product="pyrroloquinoline quinone biosynthesis protein D"
                /protein_id="BAC76464.1"
                /db_xref="GI:30698351"
                /translation="MTGLPEPTVPRLRPGVRLTRDPARGELALLPERVVVLNDTAAAV
                LAHCDGTTSLAGIVERLAEEYEGVSAEDVRELLLRLAQRRVVDLHG"
     CDS        complement(7171. .7899)
                /note="ORF7 (242 aa), lankacidin biosynthesis protein
                similar to AE004625-5 Pseudomonas aeruginosa
                pyrroloquinoline quinone biosynthesis protein C (250 aa)"
                /codon_start=1
                /transl_table=11
                /product="pyrroloquinoline quinone biosynthesis protein C"
                /protein_id="BAC76465.1"
                /db_xref="GI:30698352"
                /translation="MSMSVTREVAAPWSEAEFRQRLHALESSYWDRHPFHRRMHEGLL
                DEGELRLWAANRWYYQRCLPQKDAAIVANCPLPEVRRQWLSRIVYHDGADACAGGAEK
                WLRLAEAVGLRRDEVHDERLVLAGTRFAVDAYVDFARRRPWLEAAASGLTELFSPGLL
                AHRLGRLREHYPWIAEEGFEYFTARIEVVGPEGRSLLDLVARHAVSREQQEACVRALA
                FKCRVLNAVLDSLDYHTGNGATRS"
     CDS        complement(7926. .8030)
                /note="ORF8 (34 aa), lankacidin biosynthesis protein
                similar to AL603642-197 Sinorhizobium meliloti putative
                pyrroloquinoline quinone biosynthesis protein A (31 aa)"
                /codon_start=1
                /transl_table=11
                /product="pyrroloquinoline quinone biosynthesis protein A"
                /protein_id="BAC76466.1"
                /db_xref="GI:30698353"
                /translation="MRTSGKELPAKKAWHRPDFVTIDTGMEVTAYFSR"
     CDS        8420. .10057
                /note="ORF9 (545 aa), possible lankacidin resistance
                protein
                similar to AL133424-19 Streptomyces coelicolor putative
                ABC transporter ATP-binding protein (544 aa)"
                /codon_start=1

  Query Match             17.6%;  Score 219.2;  DB 15;  Length 210614;
  Best Local Similarity   52.9%;  Pred. No. 6.7e-41;
  Matches  616;  Conservative    0;  Mismatches  503;  Indels   45;  Gaps    5;

Qy          6 CCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATT 65
              || |  || |   ||| |||| || || || ||  | |  |  || |||||   ||| ||
Db      86558 CCGGGCCGACCGGCTGCCGTCGCTGACGGGGCTGAGGTTCTTCGCCGCGCTCCTGGTGTT 86617

Qy         66 CGCCTGTCA---CATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACT 122
              || ||| ||        ||||     |      ||||||        | | | |
Db      86618 CGTCTGCCATGCGTGGAGCCAGTTCAACCCGCAGGCCGACACCGCCTCGCGAAACCTGGC 86677

Qy        123 GCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATT 182
                | |    |||  |||  || | | ||||||  |||||||| | || || || ||
Db      86678 CTACTTCCTGCTGCCGGCCGGCCCCGTCGGGGTCTCCTTCTTCTTCGTCCTCAGCGGGTT 86737

Qy        183 CGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCAT 242
              ||| ||  | |   |||| ||    |  ||  |||| |      ||||||||||| || ||
Db      86738 CGTGCTGACGTACGCGGCGCGGCCGGGCGATACCGTTCTCGGCTTCTGGCGGCGGCGGAT 86797

Qy        243 CGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTC 302
              |||||||||||| |||   ||| ||| |||||   |  | || | |||| || |  ||
Db      86798 GGCGAAGATCTACCCGAACCATGTCGTGACGT---GGGTGGCCGGTCTCGTCCTGGTCCT 86854

Qy        303 CCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCT 362
              |   |||   | |  ||  | | |||  |           |  ||   ||    |  |
Db      86855 CGCCACGGGCGCCGCCCTGTCGGCCGGGAC-----------GACGGCCGCCGCCCTCTT 86902

Qy        363 GCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTG 422
              |||||||| | |||   ||  ||  || | | || ||||   |||| ||  |   ||
Db      86903 CCTCGTGCAGGCGTGGTGGCCGCACATGATCATGACCGCCACGGTCAACCCGATCTCGTG 86962

Qy        423 GTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCG 482
              |||||||  |||| ||| |   |||||| |||  |||||   |   |  ||||   |||||
Db      86963 GTCGCTCGCCTGCGAGCTGTTCTTCTACCTGCTGTTCCCCTTCCTGCTCCGGGCCGTGCG 87022

Qy        483 CAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTG 542
              | || | ||   | | ||||||||  |  ||| |    | |  | | |    |   |
Db      87023 CCGGCTGCCGCGGCAGGGGCTGTGGCCGGCCGCGGGCCTGCTGACGGCGCTGATCCTGGC 87082

Qy        543 TGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCT 602
              || || | ||||||    |||||  || |  |  |    |  |    | |||      ||
Db      87083 CGTACCGCAGCTGGCCACTCTGCTGCCCGGCAAGGTCATGCCGCTGGCGGGGTACAGCCT 87142

Qy        603 CAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGG 662
              ||   || |||||||||  |||      || ||||||||| ||||||||||||  |||||
Db      87143 CACGCAGTTCTGGTTCGTCTACACCGCGCCCCCGGTGCGCCTGCTGGAGTTCGCGCTCGG 87202

Qy        663 CATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTG 722
              ||||||  | || ||  |  |||||   || || |||     |  | |       | |||
Db      87203 GATCGTGCTCGCTCTCCTTGTGCGCAACGGCATGTGGCCCCTCGTCAGCGTCCGGCTCGC 87262

Qy        723 CACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGAT-----------------GGT 764
              |   |  ||   |  |||  | ||||   |  ||    |                 | |
Db      87263 CGTCGTCCTGTGCTGCGCCGGGTACGCGGTGGCCGGTGTCACGACGATGGTCACCAAGCT 87322

Qy        765 GCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGC 824
              |||||   | |   |  | ||   |  | ||||| | |||   ||| | ||| ||  |
Db      87323 GCCCCACCTGTGGGCGATGGTGAGCGCCACCGTCATCCCGCTGGCGGTCCTCATCGCCAC 87382

Qy        825 GCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCG 884
              |   |||  |  ||||  |  ||||| || | | |     ||     ||   ||| | ||
Db      87383 GGCCGCCTCCCGCGACCGGAGCGGCCTGCCCTCCGCCGCCCGCCATCCGCGCCTGATCCG 87442

Qy        885 GCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACA 944
              ||||||| |   |||||||||||||||||||| |||  | ||||| |      | |
Db      87443 GCTGGGCGAACTCTCCTTCGCCTTCTACCTGGTGCACCACATGATCCTGACGGCCTG--- 87499

Qy        945 CCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGC 1004
                    |   || ||  |||     |||      | |  | |     |  |  ||
Db      87500 ------CTTCACCGCCGTGCTGCCGGACGGCTTCCGGCCGGGTCTGCTCGGGCTCACCGG 87553

Qy       1005 GCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGT 1064
              || || | | | || |    ||          ||| | | | || ||||||||||  ||
Db      87554 CGTGGCCGTCGCGCTGTTCTTCGTCTGCTGGGCGGCGGCCTGGCTCCTGCACACCGCGGT 87613

Qy       1065 CGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGG 1124
              |||    |||  ||||||||  |  ||| |||    | |  || ||||  ||   | |
Db      87614 CGAGACACCCGCCATGCGCCGCTGGGGCCGCCCTCCGGCGACCGCCCGCGCGCCCCGTGA 87673

Qy       1125 CGCCACCGCGGCTCCCCGGAACTC 1148
                | ||   ||  || ||   ||
Db      87674 GCCGGCCCGCGCATCCGGGCGATC 87697
```

Search completed: August 25, 2006, 23:54:51
Job time : 7447 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2006, 22:01:28 ; Search time 333 Seconds
(without alignments)
6093.597 Million cell updates/sec

Title: US-10-680-860A-13
Perfect score: 1245
Sequence: 1 gtgctccagcgcgtcgatct.....aacccaccccgagaggatga 1245

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2239192 seqs, 814926892 residues

Total number of hits satisfying chosen parameters: 4478384

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09_NEW_PUB.seq:*
2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10_NEW_PUB.seq:*
7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11_NEW_PUB.seq:*
8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11_NEW_PUB.seq1:*
9: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
10: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	794	63.8	84428	8	US-11-330-363-1	Sequence 1, Appli
	2	356.4	28.6	84428	8	US-11-330-363-1	Sequence 1, Appli
	3	70.6	5.7	2133	9	US-11-218-305-13664	Sequence 13664, A
c	4	65.8	5.3	1381	6	US-10-374-780A-1512	Sequence 1512, Ap
	5	64.6	5.2	1257	6	US-10-449-902-21220	Sequence 21220, A
	6	62.6	5.0	2097	6	US-10-449-902-24239	Sequence 24239, A
	7	62.4	5.0	2052	9	US-11-056-355B-66165	Sequence 66165, A
	8	61.2	4.9	1155	6	US-10-449-902-3207	Sequence 3207, Ap
	9	61.2	4.9	1198	6	US-10-449-902-8767	Sequence 8767, Ap
	10	60.2	4.8	2062	9	US-11-174-307B-875	Sequence 875, App
	11	60.2	4.8	2062	9	US-11-056-355B-2961	Sequence 2961, Ap
	12	59.8	4.8	938	6	US-10-449-902-4345	Sequence 4345, Ap
	13	59.8	4.8	994	6	US-10-449-902-14734	Sequence 14734, A
	14	59.6	4.8	1487	9	US-11-174-307B-4517	Sequence 4517, Ap
c	15	59.2	4.8	2219	9	US-11-218-305-17428	Sequence 17428, A
c	16	59.2	4.8	5453	9	US-11-218-305-17434	Sequence 17434, A
	17	59	4.7	1329	6	US-10-449-902-19828	Sequence 19828, A
	18	59	4.7	1543	9	US-11-218-305-14842	Sequence 14842, A
	19	58.8	4.7	519	6	US-10-501-187-242	Sequence 242, App
	20	58.8	4.7	552	6	US-10-501-187-241	Sequence 241, App
	21	58.6	4.7	1225	9	US-11-056-355B-62822	Sequence 62822, A
	22	58.4	4.7	1477	9	US-11-174-307B-3397	Sequence 3397, Ap
	23	58.2	4.7	1420	9	US-11-218-305-11878	Sequence 11878, A
	24	58	4.7	9762	7	US-11-330-123-73	Sequence 73, Appl
c	25	57.8	4.6	2312	9	US-11-218-305-5837	Sequence 5837, Ap
	26	57.8	4.6	3561	9	US-11-218-305-9070	Sequence 9070, Ap
	27	57.6	4.6	1431	6	US-10-449-902-8716	Sequence 8716, Ap
	28	57.4	4.6	1319	6	US-10-449-902-3354	Sequence 3354, Ap
	29	57.4	4.6	2097	9	US-11-218-305-11877	Sequence 11877, A
	30	57.2	4.6	945	8	US-11-266-748A-114634	Sequence 114634,
c	31	57.2	4.6	945	8	US-11-266-748A-120694	Sequence 120694,
	32	57.2	4.6	945	8	US-11-266-748A-191155	Sequence 191155,
	33	57.2	4.6	945	8	US-11-266-748A-225659	Sequence 225659,
	34	57.2	4.6	1011	8	US-11-266-748A-165427	Sequence 165427,
	35	57.2	4.6	1073	8	US-11-266-748A-348600	Sequence 348600,
	36	57.2	4.6	1073	8	US-11-266-748A-381914	Sequence 381914,
c	37	57.2	4.6	1073	8	US-11-266-748A-431979	Sequence 431979,
	38	57.2	4.6	1093	8	US-11-266-748A-74860	Sequence 74860, A
	39	57.2	4.6	1093	8	US-11-266-748A-108601	Sequence 108601,
c	40	57.2	4.6	1093	8	US-11-266-748A-127671	Sequence 127671,
	41	57.2	4.6	1595	8	US-11-266-748A-191154	Sequence 191154,
	42	57.2	4.6	1595	8	US-11-266-748A-225658	Sequence 225658,
	43	57.2	4.6	1612	8	US-11-266-748A-348602	Sequence 348602,
	44	57.2	4.6	1612	8	US-11-266-748A-381916	Sequence 381916,
c	45	57.2	4.6	1612	8	US-11-266-748A-431981	Sequence 431981,

ALIGNMENTS

RESULT 1
US-11-330-363-1/c
; Sequence 1, Application US/11330363
; Publication No. US20060121577A1
; GENERAL INFORMATION:
; APPLICANT: Meiji Seika Kaisha, Ltd.
; TITLE OF INVENTION: Midecamycin biosynthetic genes
; FILE REFERENCE: 138451 US
; CURRENT APPLICATION NUMBER: US/11/330,363
; CURRENT FILING DATE: 2006-01-12
; PRIOR APPLICATION NUMBER: US/10/229,148
; PRIOR FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 210516/2002
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 84428
; TYPE: DNA
; ORGANISM: Streptomyces mycarofaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((1)..(675))
; OTHER INFORMATION: ORF42 (fragment)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((1168)..(2202))
; OTHER INFORMATION: ORF41
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((2220)..(3215))
; OTHER INFORMATION: ORF40
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((3237)..(4691))
; OTHER INFORMATION: ORF39
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((4695)..(5948))
; OTHER INFORMATION: ORF38
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6048)..(6629))
; OTHER INFORMATION: ORF37
; FEATURE:
; NAME/KEY: CDS

; LOCATION: Complement((6653)..(7945))
; OTHER INFORMATION: ORF36
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8149)..(9015)
; OTHER INFORMATION: ORF35
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9012)..(9335)
; OTHER INFORMATION: ORF34
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9328)..(10458)
; OTHER INFORMATION: ORF33
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10521)..(11603)
; OTHER INFORMATION: ORF32
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11729)..(12961)
; OTHER INFORMATION: ORF31
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((13016)..(14044))
; OTHER INFORMATION: ORF30
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((14074)..(15096))
; OTHER INFORMATION: ORF29
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15643)..(17466)
; OTHER INFORMATION: ORF28
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((17522)..(18895))
; OTHER INFORMATION: ORF27
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19063)..(20229)
; OTHER INFORMATION: ORF26
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((20307)..(21743))
; OTHER INFORMATION: ORF25
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((21733)..(22527))
; OTHER INFORMATION: ORF24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((22534)..(23571))
; OTHER INFORMATION: ORF23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((23555)..(24463))
; OTHER INFORMATION: ORF22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((24460)..(25650))
; OTHER INFORMATION: ORF21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((25647)..(26105))
; OTHER INFORMATION: ORF20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((26180)..(27391))
; OTHER INFORMATION: ORF19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27937)..(28983)
; OTHER INFORMATION: ORF18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29244)..(42779)
; OTHER INFORMATION: ORF1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42823)..(48657)
; OTHER INFORMATION: ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (48712)..(59802)
; OTHER INFORMATION: ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59850)..(64556)
; OTHER INFORMATION: ORF4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64687)..(70365)
; OTHER INFORMATION: ORF5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70365)..(71078)
; OTHER INFORMATION: ORF6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71113)..(72360)
; OTHER INFORMATION: ORF7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72400)..(73665)
; OTHER INFORMATION: ORF8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73694)..(75043)
; OTHER INFORMATION: ORF9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((75899)..(76570))
; OTHER INFORMATION: ORF10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((76602)..(77765))
; OTHER INFORMATION: ORF11
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78039)..(79313)
; OTHER INFORMATION: ORF12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((79391)..(81052))
; OTHER INFORMATION: ORF13
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81541)..(82356)
; OTHER INFORMATION: ORF14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82760)..(83362)
; OTHER INFORMATION: ORF15
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((83495)..(84142))
; OTHER INFORMATION: ORF16
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (84329)..(84428)
; OTHER INFORMATION: ORF17 (fragment)

US-11-330-363-1

Query Match 63.8%; Score 794; DB 8; Length 84428;
Best Local Similarity 80.9%; Pred. No. 8.3e-137;

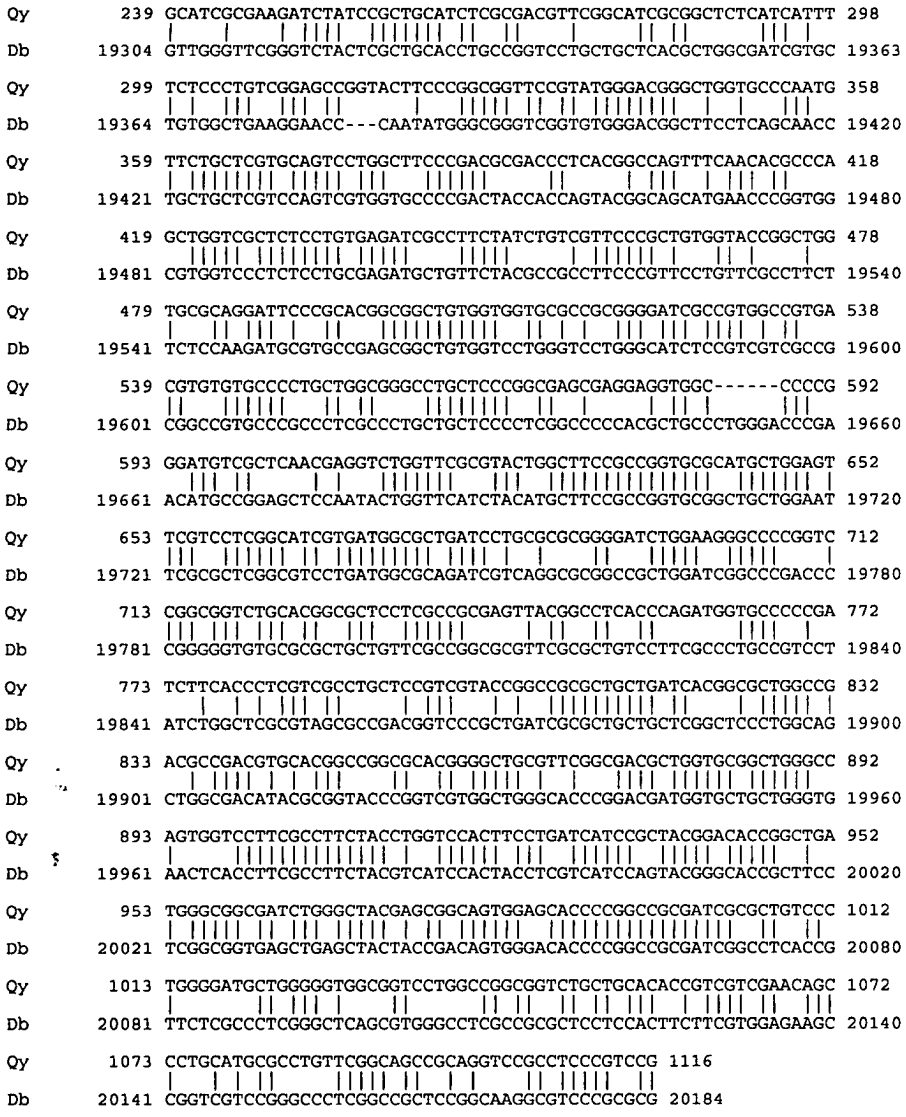
```
  Matches  937;  Conservative    0;  Mismatches  220;  Indels    1;  Gaps    1;

Qy         12 CGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTG 71
              ||||   ||| |||| |||||||||||| |||||||  |||||  |  ||||||||||||||
Db      77751 CGTCCGCCTGCCGTCGCTCACCGGCCTTCGCTGGTTCGCGGCCTTAGCGGTATTCGCCTG 77692

Qy         72 TCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACAT 131
              |||||| |||||||||    |||||||| |||||||||||| || | ||| ||||| |||||
Db      77691 TCACATTGCCCAGCAGCAATTCTTCGCTGACCAGCAGGTCGGAACCGCGCTGCTACACAT 77632

Qy        132 CACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGC 191
              |||| |||||||||| || ||||||||| |||||||| | || || || || || || || ||
Db      77631 CACCACGCTCGGTTCGATCGCGGTCTCGGTCTTCTTCCTGCTCAGCGGTTTTGTTCTGGC 77572

Qy        192 CTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGAT 251
              |||||||||| ||| ||||||| ||  ||||||||||||||||||  | || |||||
Db      77571 GTGGTCGGCCCGTGACAAGGACTCTGTCACGACTTTCTGGCGGCGCCGATTTGCCAAGAT 77512

Qy        252 CTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGA 311
              ||| |||||||| |||| || ||| ||| |||| ||||||||||||||  ||||
Db      77511 CTACCCGCTGCACCTCGTAACTTTCCTCATAGCGGGCGTCATCATTTTCTCCCTCGCGGA 77452

Qy        312 GCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCA 371
              ||||   || ||||| |||||||||||||||| ||||| ||| || ||||||| |||||
Db      77451 GCCGACTCTGCCCGGTGGTTCCGTATGGGACGGTCTGGTTCCCGATCTTCTGCTGGTGCA 77392

Qy        372 GTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTC 431
              |||||||||||||| | ||  | |  ||| | |||||||| || || ||||||||||
Db      77391 GTCCTGGCTTCCCGAACCCACGATTATCGCCGGGTTCAACACCCCGAGTTGGTCGCTCTC 77332

Qy        432 CTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCC 491
              |||||| |||| ||||||||| ||||||||||||||| |||||||||| | ||||||
Db      77331 CTGTGAATTCGCTTTCTATCTGACGTTCCCGCTGTGGTATCGGCTGGTGCGGAAGATTCC 77272

Qy        492 CGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCT 551
              | |||||||||||||||||| ||||| || || |||| |||||||  ||||| ||  |
Db      77271 GGTACGGCGGCTGTGGTGGTGTGCCGCCGGTATTGCCGCGGCCGTGATTTGTGTACCGTT 77212

Qy        552 GCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGT 611
              || || ||| | |||||||||||  |||  |||||||||||| |||||||||| |
Db      77211 CGTGACGAGCCAGTTCCCGGCGAGCGCCGAGACGGCCCCCGGGATGCCGCTCAACGAGCT 77152

Qy        612 CTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGAT 671
              |||||||| | |||||| |||||||||| ||||||||||||||||||||||||||||||||
Db      77151 GTGGTTCGCCTGCTGGCTGCCGCCGGTGCGGATGCTGGAGTTCGTCCTCGGCATCGTGAT 77092

Qy        672 GGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCT 731
              |||||||||||||||| ||||  | |||  |||||||||   || |  || | ||  |||
Db      77091 GGCGCTGATCCTGCGCACGGGCGTGTGGCGGGGCCCCGGAGTGGTGTCCTCCGCGCTGCT 77032

Qy        732 CCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTG 791
              |||||||||   |||||| |||||||| |||||||||||| |||||| ||| ||| ||
Db      77031 GCTCGCCGCGGCCTACGGCGTCACCCAGGTGGTGCCCCCGATGTTCACCATCGCCGCGTG 76972

Qy        792 CTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCG 851
              |||| |||| || ||||| |||||||||| ||  ||||| | ||||| ||||| ||||
Db      76971 CTCCATCGTGCCCGCCGCCCTGCTGATCACCGCTTTGGCCAATGCCGATGTGCAGGGCCT 76912

Qy        852 GCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTA 911
              ||||||  ||  |||| ||||||   |||||||||||||| ||||||||||||||||||
Db      76911 GCGCACCGGTTTGCGCTCGGCGGTACTGGTGCGGCTGGGCGAGTGGTCCTTCGCCTTCTA 76852

Qy        912 CCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTA 971
              ||||| |||||| || |||||||||||| |||||||||||||||||||| ||||||||
Db      76851 TCTGGTTCACTTCATGGTCATCCGCTACGGGCACCGGCTGATGGGCGGCGAACTGGGCTA 76792

Qy        972 CGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGC 1031
              || ||| |||||||||||| ||  ||||  |||||| |||||| |||||||| ||| ||
Db      76791 CGCGCGCCAGTGGAGCACCGCGAGCGCGGGAGCGCTGGCCCTGGCGATGCTGGCGGTCGC 76732

Qy       1032 GGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGG 1091
              | || ||||||| || |||||||||||||||| ||  | |||||||||| ||| ||||
Db      76731 GATCGTGGCCGGTGGGCTGCTGCACACCGTCGTGGAGAACCCCTGCATGCGTCTGCTCGG 76672

Qy       1092 CAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG-CGGCTCCCCGGAACTCAC 1150
              | ||||| |  | | | ||    |       |  ||| |||| ||   | ||| |||||||
Db      76671 CCGCCGCCGCCCGGTCGCCACCGCACCAGACCCCGCAACCGACGAGGCTCCGAAACTCAC 76612

Qy       1151 CCGCGGCCGACGCGGCCG 1168
              ||||| | ||||   |||
Db      76611 CCGCGCCTGACGTCACCG 76594


RESULT 2
US-11-330-363-1
; Sequence 1, Application US/11330363
; Publication No. US20060121577A1
; GENERAL INFORMATION:
;  APPLICANT: Meiji Seika Kaisha, Ltd.
;  TITLE OF INVENTION: Midecamycin biosynthetic genes
;  FILE REFERENCE: 138451 US
;  CURRENT APPLICATION NUMBER: US/11/330,363
;  CURRENT FILING DATE:  2006-01-12
;  PRIOR APPLICATION NUMBER: US/10/229,148
;  PRIOR FILING DATE: 2002-08-28
;  PRIOR APPLICATION NUMBER: 210516/2002
;  PRIOR FILING DATE: 2002-07-19
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 84428
;   TYPE: DNA
;   ORGANISM: Streptomyces mycarofaciens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((1)..(675))
;   OTHER INFORMATION: ORF42 (fragment)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((1168)..(2202))
;   OTHER INFORMATION: ORF41
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((2220)..(3215))
;   OTHER INFORMATION: ORF40
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((3237)..(4691))
;   OTHER INFORMATION: ORF39
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((4695)..(5948))
;   OTHER INFORMATION: ORF38
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((6048)..(6629))
;   OTHER INFORMATION: ORF37
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((6653)..(7945))
;   OTHER INFORMATION: ORF36
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (8149)..(9015)
;   OTHER INFORMATION: ORF35
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (9012)..(9335)
;   OTHER INFORMATION: ORF34
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (9328)..(10458)
;   OTHER INFORMATION: ORF33
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10521)..(11603)
```

; OTHER INFORMATION: ORF32
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11729)..(12961)
; OTHER INFORMATION: ORF31
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((13016)..(14044))
; OTHER INFORMATION: ORF30
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((14074)..(15096))
; OTHER INFORMATION: ORF29
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15643)..(17466)
; OTHER INFORMATION: ORF28
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((17522)..(18895))
; OTHER INFORMATION: ORF27
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19063)..(20229)
; OTHER INFORMATION: ORF26
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((20307)..(21743))
; OTHER INFORMATION: ORF25
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((21733)..(22527))
; OTHER INFORMATION: ORF24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((22534)..(23571))
; OTHER INFORMATION: ORF23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((23555)..(24463))
; OTHER INFORMATION: ORF22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((24460)..(25650))
; OTHER INFORMATION: ORF21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((25647)..(26105))
; OTHER INFORMATION: ORF20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((26180)..(27391))
; OTHER INFORMATION: ORF19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27937)..(28983)
; OTHER INFORMATION: ORF18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29244)..(42779)
; OTHER INFORMATION: ORF1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42823)..(48657)
; OTHER INFORMATION: ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (48712)..(59802)
; OTHER INFORMATION: ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59850)..(64556)
; OTHER INFORMATION: ORF4

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64687)..(70365)
; OTHER INFORMATION: ORF5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70365)..(71078)
; OTHER INFORMATION: ORF6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71113)..(72360)
; OTHER INFORMATION: ORF7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72400)..(73665)
; OTHER INFORMATION: ORF8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73694)..(75043)
; OTHER INFORMATION: ORF9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((75899)..(76570))
; OTHER INFORMATION: ORF10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((76602)..(77765))
; OTHER INFORMATION: ORF11
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78039)..(79313)
; OTHER INFORMATION: ORF12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((79391)..(81052))
; OTHER INFORMATION: ORF13
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81541)..(82356)
; OTHER INFORMATION: ORF14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82760)..(83362)
; OTHER INFORMATION: ORF15
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((83495)..(84142))
; OTHER INFORMATION: ORF16
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (84329)..(84428)
; OTHER INFORMATION: ORF17 (fragment)
US-11-330-363-1

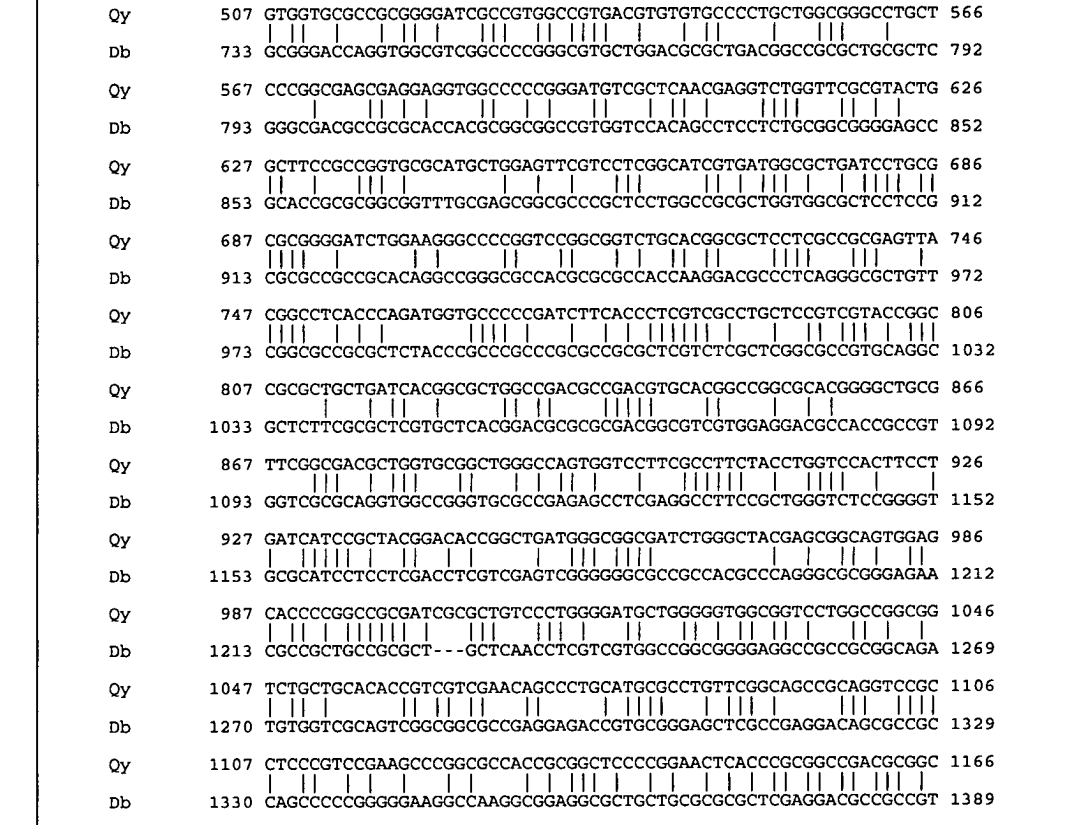
Query Match 28.6%; Score 356.4; DB 8; Length 84428;
Best Local Similarity 59.3%; Pred. No. 1.3e-56;
Matches 666; Conservative 0; Mismatches 446; Indels 12; Gaps 3;

```
Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61
              ||| | ||      | ||| |||| |||  |||| | ||       || || |||  ||
Db      19064 TGCCCTTGCCGAAACACCTGCCGTCGCTCGGCGGCATGCGGGCCATCGCCGCACTGGTGG 19123

Qy         62 TATTCGCCTGTCACATCG---CCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCG 118
              | |||  || ||| ||||   |||||| |   |||  |  || | ||       || ||
Db      19124 TGTTCTGCTCTCATATCGCTTCCCAGCCGTTTTTCCGCAACGCCAAGATAAACTCCACCG 19183

Qy        119 CACTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTG 178
              ||| | | |   |   |   || ||  | |||||||||| |||||||||| || || |
Db      19184 CACAGGTCCCGCTGGACGTCCTGGGGCCGCTGGCGGTCTCGTTCTTCTTCATGCTCAGCG 19243

Qy        179 GATTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCC 238
              |||||||||| ||||| ||| |     |||    |||  |||  | |||||||| ||||
Db      19244 GATTCGTCCTCACCTGGGCGGGCATGCCCGACCCGTCCAAGCCTGCCTTCTGGCGCCGCC 19303
```



RESULT 3
US-11-218-305-13664
; Sequence 13664, Application US/11218305
; Publication No. US20060141495A1
; GENERAL INFORMATION:
; APPLICANT: MONSANTO TECHNOLOGY, LLC
; APPLICANT: McLaird, Paul L.
; APPLICANT: Tao, Nengbing
; APPLICANT: Wu, Kunsheng
; TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping
; TITLE OF INVENTION: Corn.
; FILE REFERENCE: 38-21 (53660)B
; CURRENT APPLICATION NUMBER: US/11/218,305
; CURRENT FILING DATE: 2005-09-01
; PRIOR APPLICATION NUMBER: US 60/606,880
; PRIOR FILING DATE: 2004-09-01
; NUMBER OF SEQ ID NOS: 25043
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13664
; LENGTH: 2133
; TYPE: DNA
; ORGANISM: Zea mays
US-11-218-305-13664

Query Match 5.7%; Score 70.6; DB 9; Length 2133;
Best Local Similarity 45.1%; Pred. No. 0.00024;
Matches 305; Conservative 0; Mismatches 369; Indels 3; Gaps 1;



RESULT 4
US-10-374-780A-1512/c
; Sequence 1512, Application US/10374780A
; Publication No. US20060162006A9
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis

```
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Heard, Jacqueline E
;  APPLICANT:  Haake, Volker
;  APPLICANT:  Creelman, Robert A
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Broun, Pierre E
;  APPLICANT:  Pilgrim, Marsha L
;  APPLICANT:  Dubell III, Arnold T
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Yu, Guo-Liang
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
;  FILE REFERENCE: MBI-0047 CIP
;  CURRENT APPLICATION NUMBER: US/10/374,780A
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: 09/837,944
;  PRIOR FILING DATE: 2001-04-18
;  PRIOR APPLICATION NUMBER: 60/310,847
;  PRIOR FILING DATE: 2001-08-09
;  PRIOR APPLICATION NUMBER: 09/934,455
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/336,049
;  PRIOR FILING DATE: 2001-11-19
;  PRIOR APPLICATION NUMBER: 60/338,692
;  PRIOR FILING DATE: 2001-12-11
;  PRIOR APPLICATION NUMBER: 10/171,468
;  PRIOR FILING DATE: 2002-06-14
;  PRIOR APPLICATION NUMBER: 10/225,066
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: 10/225,067
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: 10/225,068
;  PRIOR FILING DATE: 2002-08-09
;  NUMBER OF SEQ ID NOS: 2906
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1512
;   LENGTH: 1381
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G1255
US-10-374-780A-1512

  Query Match             5.3%;  Score 65.8;  DB 6;  Length 1381;
  Best Local Similarity   45.5%;  Pred. No. 0.0017;
  Matches  270;  Conservative    0;  Mismatches  322;  Indels    1;  Gaps    1;

Qy        597 GTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGT 656
              | ||| |  || ||    |  |||      |   |||  ||  | || |  ||| |
Db        852 GACGCACGCCGTGGCGACGGCGGCGGCGGCGGCGGCGCGCTTGACGAAGCGGCCCTTCAT 793

Qy        657 CCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGC 716
              || ||||  | |   |||| ||| | ||||| |   |  | |||    |  |    |
Db        792 CCGCGGCCGCTTCTCGGCGTTGAGCTTGCGCACCTCGTACCGGATCTTCTTGGAGAAGAG 733

Qy        717 GGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTT 776
                |  |      ||  |     |  ||  || | ||    ||  |||  | |||
Db        732 CCGCGTCCGCCGCTTCTCCCTGTACCGCGACACCCGCGCCTCCCTGCCGCCGTCCATCCC 673

Qy        777 CACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGC 836
              || ||||| || | || ||  |    || |  ||   | || |||  |    ||||  |
Db        672 CAGCCTCGGCGTCAGCGCCTGCTCCTCGCCGCCGTGCCCGAACACCCCTGCCGCCGCCAT 613

Qy        837 CGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTG 896
              | ||   | ||     |     ||| | |  ||| |||   ||  |||     ||  |
Db        612 CCACACCCCCGAGTAGTCGCGGGGCCAGGAGTCGTCGAGCTTGACGCGCGGCCGCTCGCC 553

Qy        897 GTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGG 956
              |||  ||  | | |  |  ||| ||  ||| ||  ||           ||||||  ||
Db        552 GTCGGTCCACGGCGAGGTCCCCCAGCTCTCGATGATGGCCTCGTAGTTGAGGCTGAGGGC 493

Qy        957 CGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGG 1016
              |    ||  |||   |  ||  || |     |  |||| |||   |  | |   ||
Db        492 GAGGCTGCTCTGCTGGAACTGGGCGTCGTCGTCGGCGGCGGCGGCGGGCCCGCTGCTCCC 433

Qy       1017 GATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTG 1076
              ||   |  |  ||  ||| | |||| | |     ||                 | || |
Db        432 GACCTTCTCGTCGGGTGTCGTCGCCGCCTGAGGTGAGCCGTAGTCGAAGTCGATGTCCAG 373

Qy       1077 CATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGC 1136
              |||    | |  || |  |  |  |||| |||  | |   |  |  | | || | ||||
Db        372 CATCTCGCCGGACGACGTCTCCGTGTCCAGCTCGTGGCACCACGCTAGGGTCATCCCGGC 313

Qy       1137 TCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCG 1189
              | | |    |  || |    ||| || |||||||   | | | || |  ||
Db        312 GTCGCCGTCGGCCTCCAC-CTTGACCCGCCCGGCGTCGTCGTCCACCGGCTCG 261


RESULT 5
US-10-449-902-21220
; Sequence 21220, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
;  APPLICANT: National Institute of Agrobiological Sciences.
;  APPLICANT:  Bio-oriented Technology Research Advancement Institution.
;  APPLICANT:  The Institute of Physical and Chemical Research.
;  APPLICANT:  Foundation for Advancement of International Science.
;  TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
;  FILE REFERENCE: MOA-A0205Y1-US
;  CURRENT APPLICATION NUMBER: US/10/449,902
;  CURRENT FILING DATE:  2003-05-29
;  PRIOR APPLICATION NUMBER: JP 2002-203269
;  PRIOR FILING DATE: 2002-05-30
;  PRIOR APPLICATION NUMBER: JP 2002-383870
;  PRIOR FILING DATE: 2002-12-11
;  NUMBER OF SEQ ID NOS: 56791
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21220
;   LENGTH: 1257
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: AK099178
;   DATABASE ENTRY DATE: 2002-08-28
US-10-449-902-21220

  Query Match             5.2%;  Score 64.6;  DB 6;  Length 1257;
  Best Local Similarity   47.4%;  Pred. No. 0.0029;
  Matches  193;  Conservative    0;  Mismatches  214;  Indels    0;  Gaps    0;

Qy        642 CATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAA 701
              ||   ||  |  || || || | ||     |  |||  ||||   ||  | | || |
Db        107 CACAATGCCGCGCGCCCGCGTCGTCCACGTCATGCTCCTCCTCTTCGTCGCGCTCGGCGC 166

Qy        702 GGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGAT 761
              | |||  |   |       |  ||| ||  |||||||||        || |||| || || |
Db        167 GTGCCGTGCCACCCACAACATCACCGCCATCCTCGCCGACCACCGCGACCTCGCCGAGTT 226

Qy        762 GGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCAC 821
              |  | || | ||     ||   | |  |      || | |  |             |||||
Db        227 CGGCCGCCAGCTCACGGCCACGGGCCTCGCGGACGACATCGACGGGCGCAACACCATCAC 286

Qy        822 GGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGT 881
              || ||| |||| || |||||  |||     ||      ||  |||   |   |||  |
Db        287 GGTGCTCGCCGTCGACGACGCCCACATGGCGCAGCTCAGGGCGCGCGGGCTCCCGCGCGA 346

Qy        882 GCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGG 941
              |  |||| ||||   |  ||       |  ||| ||| ||| |     |  |  | |
Db        347 GGCGCTGCGCCACGTGCTCTCGCTGCACGTCCTCGTCGACTACTACGACGACGCCAAGCT 406

Qy        942 ACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGAT 1001
              |||||  ||    ||||||    || |       || ||       |  ||| ||
Db        407 GCACCGCCTCCCCGGCGGCTCCGCGGTCGTGTCCACGCTGTTCCAGGCGTCGGGCGACGC 466
```

```
Qy       1002 CGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTC 1048
Db        467 GCCCGGGAGCGAGGGGATGGTGAAGATCGCCGTGCGCCGCGGCGGCC 513
```

RESULT 6
US-10-449-902-24239
; Sequence 24239, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24239
; LENGTH: 2097
; TYPE: DNA
; ORGANISM: Oryza sativa
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK099681
; DATABASE ENTRY DATE: 2002-08-28
US-10-449-902-24239

Query Match 5.0%; Score 62.6; DB 6; Length 2097;
Best Local Similarity 49.2%; Pred. No. 0.0069;
Matches 164; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

```
Qy        912 CCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTA 971
Db         31 CCTCACCCACCACCCCACCCATCGCCTCCGCCACCGGCGAGAGGAGGCGGAGCTGCGCGT 90

Qy        972 CGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGC 1031
Db         91 GGCGATGAAGGGGTCCCACTCGCAGTCGCAGGCCCAGGCCCAGTCCCAGGCGGGCCGGCG 150

Qy       1032 GGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGG 1091
Db        151 GCGGCGGTGCGGGTGGCTGCTGCCGCTCCTCGTCGGCGCGGCCTTCCTCGCCGAGATCGC 210

Qy       1092 CAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACC 1151
Db        211 GTTCCTCGGCCGCCTCGACATGGCGAAGAACGCCGCCGCGGTCGAGAGCTGGACCACCTC 270

Qy       1152 CGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGCGCACACCCC 1211
Db        271 CTTCTACGCCCGCTCCTCCGCCCCCGCCCGCGACGGGAAGGCAGCGGTCGTGGTCCCCGG 330

Qy       1212 CGCAGCGACGAACGAACCCACCCCGAGAGGATG 1244
Db        331 CGCGGACGCGGACGACGCGCCCCCGGGCGGAGG 363
```

RESULT 7
US-11-056-355B-66165
; Sequence 66165, Application US/11056355B
; Publication No. US20060150283A1
; GENERAL INFORMATION:
; APPLICANT: Brover, Vyacheslav
; APPLICANT: Alexandrov, Nickolai
; TITLE OF INVENTION: Sequence Determined DNA Fragments and Corresponding
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: 2750-1590PUS2
; CURRENT APPLICATION NUMBER: US/11/056,355B
; CURRENT FILING DATE: 2005-02-14
; PRIOR APPLICATION NUMBER: 60/544,190
; PRIOR FILING DATE: 2004-02-13
; NUMBER OF SEQ ID NOS: 119966
; SEQ ID NO 66165
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2052)
; OTHER INFORMATION: Ceres Seq. ID no. 13657610
US-11-056-355B-66165

Query Match 5.0%; Score 62.4; DB 9; Length 2052;
Best Local Similarity 45.7%; Pred. No. 0.0075;
Matches 265; Conservative 0; Mismatches 306; Indels 9; Gaps 1;

```
Qy        649 GAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCC 708
Db        846 GCGGGCGTCATCTCCATCGACCTGGCGCGTGACATGCTCCAGGCCAGCGGCGCTGGGCTC 905

Qy        709 GGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCC 768
Db        906 GCCGTCGTGGTGAGCACCGAGGCCGTGTCATTCACGTGGTACGCCGGGAAGCGCCGGTCC 965

Qy        769 CCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTG 828
Db        966 ATGCTCATCCCCAACGCCTTCTTCCGCGCGGGCGCCGCCGCCGTGCT---------GCTG 1016

Qy        829 GCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTG 888
Db       1017 TCGAACCGCCGCAGGGACTTCCGCCGCGCCAAGTACCAGCTGGAGCACGTGGTGCGCACG 1076

Qy        889 GGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGG 948
Db       1077 CACAAGGGCGCCGACGACCGCGCCTTCCGCTCCGTGTACCAGGAGGAGGACGAGCAGCGG 1136

Qy        949 CTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTG 1008
Db       1137 ATCAAGGGCCTGTCCATCAGCCGCGACCTGGTGGAGGTCGGCGGCCACGCGCTCAAGACC 1196

Qy       1009 TCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAA 1068
Db       1197 AACATCACCACCCTCGGCCCGCTGGTGCTGCCCTTCTCGGAGCAGCTCCTCTTCTTCGCC 1256

Qy       1069 CAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCC 1128
Db       1257 GGCGTGCTGTTCCGCCACCTCTACCCGTCCAAGACGTCCACCCCGCCGCCGCCGGCCGCC 1316

Qy       1129 ACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGC 1188
Db       1317 AACGGGGACACCTCGGCCGCCGCGCCCTACATCCCGGACTTCAAGCGCGCGTTCGAGCAC 1376

Qy       1189 GTACCCGGGCCCGCGCACACCCCCGCAGCGACGAACGAAC 1228
Db       1377 TTCTGCATGCACGCGGCCAGCCGCGACGTGCTGGAGCACC 1416
```

RESULT 8
US-10-449-902-3207
; Sequence 3207, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269

```
;  PRIOR FILING DATE: 2002-05-30
;  PRIOR APPLICATION NUMBER: JP 2002-383870
;  PRIOR FILING DATE: 2002-12-11
;  NUMBER OF SEQ ID NOS: 56791
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3207
;   LENGTH: 1155
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: AK060938
;   DATABASE ENTRY DATE: 2001-12-06
US-10-449-902-3207

  Query Match             4.9%;  Score 61.2;  DB 6;  Length 1155;
  Best Local Similarity   44.1%;  Pred. No. 0.012;
  Matches  307;  Conservative    0;  Mismatches  383;  Indels    6;  Gaps    1;

Qy        514 GCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCG 573
              || | |  ||  ||||| |    | ||    | |  ||||||| |||  ||||  || ||
Db        301 GCTGAGACGAAGGCCGTCGTGCGGCCGGAGCTCCGTCTGCTGGTGGGGGTGCTGACGACG 360

Qy        574 AGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCG 633
                 ||  |   |  |  | ||         |   |  ||  |         | |   |     |
Db        361 CCGAAGCGGTACGAGCGTCGCAACATCGTGCGCCTGGCGTACGCGCTGCAGCCGGCGGTG 420

Qy        634 CCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGG 693
              ||      || |   | || | | | || || ||||| |    |     |       | ||| |
Db        421 CCCCCCGGCGTCGCGCAGGTGGACGTCCGCTTCGTGTTCTGCCGCGTGGCCGACCCGGTG 480

Qy        694 ATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTC 753
                |  | ||      |  | || || |     |  |  |  |  |     | |  |
Db        481 GACGCGCAGCTGGTGGTGCTGGAGGCCGCGCGGCACGGCGACATCCTGGTGCTCAACTGC 540

Qy        754 ACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTG 813
              ||  |||    ||  |     |   || | ||    | ||||  ||| |||      | |
Db        541 ACGGAGAACATGAACGACGGCAAGACGCACGAGTACCTCTCCTCCGTGCCGCGGATGTTC 600

Qy        814 CTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCG 873
                |    || |    | | |||       |  || |     ||||   || ||   |||
Db        601 GCGTCGTCGCCCTACGACTACGTGATGAAGACCGACGACGACACGTACCTCCGCGTGGCC 660

Qy        874 ACGCTGGT------GCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTG 927
              |||||||      ||  | |  | ||      |  || | |||||| | | ||  | |
Db        661 GCGCTGGTCGACGAGCTCCGGCACAAGCCCCGCGACGACGTCTACCTCGGCTACGGCTTC 720

Qy        928 ATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGC 987
              | || || |||  |     || | |       ||   |||||||||    |   |||  |
Db        721 GCCGTCGGCGACGACCCGATGCAGTTCATGCACGGGATGGGCTACGTCGTGTCCTGGGAC 780

Qy        988 ACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGT 1047
              | || |   | |  ||      |  ||  || ||| ||      | |  |  ||||
Db        781 GTCGCGACGTGGGTGTCGACCAACGAGGACATCCTGCGGTACAACGACACGCACGGCCCC 840

Qy       1048 CTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCC 1107
              |   | |  |||||  ||  || |  |||  |||    ||||    | ||| |  |
Db        841 GAGGACCTCCTCGTCGGGAAATGGCTCAACATCGGCCGCCGGGGCAAGAACCGGTACAGC 900

Qy       1108 TCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCC 1167
              | | |||  |     || | |       |  | ||      | | | ||| | | |
Db        901 CTCAGGCCGCGGATGTACGACCTCAACTGGGACATGGACAACTTCCGGCCGGACACCGTC 960

Qy       1168 GGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGCG 1203
              |||||  | |||||  || |   | | ||   |||
Db        961 CTCGTGCACATGCTCAAGGACAACCGCCGGTGGGCG 996


RESULT 9
US-10-449-902-8767
; Sequence 8767, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
;  APPLICANT: National Institute of Agrobiological Sciences.
;  APPLICANT:  Bio-oriented Technology Research Advancement Institution.
;  APPLICANT:  The Institute of Physical and Chemical Research.
;  APPLICANT:  Foundation for Advancement of International Science.
;  TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
;  FILE REFERENCE: MOA-A0205Y1-US
;  CURRENT APPLICATION NUMBER: US/10/449,902
;  CURRENT FILING DATE:  2003-05-29
;  PRIOR APPLICATION NUMBER: JP 2002-203269
;  PRIOR FILING DATE: 2002-05-30
;  PRIOR APPLICATION NUMBER: JP 2002-383870
;  PRIOR FILING DATE: 2002-12-11
;  NUMBER OF SEQ ID NOS: 56791
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8767
;   LENGTH: 1198
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: AK106115
;   DATABASE ENTRY DATE: 2002-08-28
US-10-449-902-8767

  Query Match             4.9%;  Score 61.2;  DB 6;  Length 1198;
  Best Local Similarity   44.1%;  Pred. No. 0.012;
  Matches  307;  Conservative    0;  Mismatches  383;  Indels    6;  Gaps    1;

Qy        514 GCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCG 573
              || | |  ||  ||||| |    | ||    | |  ||||||| |||  ||||  || ||
Db        301 GCTGAGACGAAGGCCGTCGTGCGGCCGGAGCTCCGTCTGCTGGTGGGGGTGCTGACGACG 360

Qy        574 AGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCG 633
                 ||  |   |  |  | ||         |   |  ||  |         | |   |     |
Db        361 CCGAAGCGGTACGAGCGTCGCAACATCGTGCGCCTGGCGTACGCGCTGCAGCCGGCGGTG 420

Qy        634 CCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGG 693
              ||      || |   | || | | | || || ||||| |    |     |       | ||| |
Db        421 CCCCCCGGCGTCGCGCAGGTGGACGTCCGCTTCGTGTTCTGCCGCGTGGCCGACCCGGTG 480

Qy        694 ATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTC 753
                |  | ||      |  | || || |     |  |  |  |  |     | |  |
Db        481 GACGCGCAGCTGGTGGTGCTGGAGGCCGCGCGGCACGGCGACATCCTGGTGCTCAACTGC 540

Qy        754 ACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTG 813
              ||  |||    ||  |     |   || | ||    | ||||  ||| |||      | |
Db        541 ACGGAGAACATGAACGACGGCAAGACGCACGAGTACCTCTCCTCCGTGCCGCGGATGTTC 600

Qy        814 CTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCG 873
                |    || |    | | |||       |  || |     ||||   || ||   |||
Db        601 GCGTCGTCGCCCTACGACTACGTGATGAAGACCGACGACGACACGTACCTCCGCGTGGCC 660

Qy        874 ACGCTGGT------GCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTG 927
              |||||||      ||  | |  | ||      |  || | |||||| | | ||  | |
Db        661 GCGCTGGTCGACGAGCTCCGGCACAAGCCCCGCGACGACGTCTACCTCGGCTACGGCTTC 720

Qy        928 ATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGC 987
              | || || |||  |     || | |       ||   |||||||||    |   |||  |
Db        721 GCCGTCGGCGACGACCCGATGCAGTTCATGCACGGGATGGGCTACGTCGTGTCCTGGGAC 780

Qy        988 ACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGT 1047
              | || |   | |  ||      |  ||  || ||| ||      | |  |  ||||
Db        781 GTCGCGACGTGGGTGTCGACCAACGAGGACATCCTGCGGTACAACGACACGCACGGCCCC 840

Qy       1048 CTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCC 1107
              |   | |  |||||  ||  || |  |||  |||    ||||    | ||| |  |
Db        841 GAGGACCTCCTCGTCGGGAAATGGCTCAACATCGGCCGCCGGGGCAAGAACCGGTACAGC 900

Qy       1108 TCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCC 1167
              | | |||  |     || | |       |  | ||      | | | ||| | | |
Db        901 CTCAGGCCGCGGATGTACGACCTCAACTGGGACATGGACAACTTCCGGCCGGACACCGTC 960

Qy       1168 GGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGCG 1203
              |||||  | |||||  || |   | | ||   |||
Db        961 CTCGTGCACATGCTCAAGGACAACCGCCGGTGGGCG 996
```

```
RESULT 10
US-11-174-307B-875
; Sequence 875, Application US/11174307B
; Publication No. US20060143729A1
; GENERAL INFORMATION:
;  APPLICANT: ALEXANDROV, Nickolai
;  APPLICANT:  BROVER, Vyacheslav
;  TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND POLYPEPTIDES ENCODED THEREBY
;  TITLE OF INVENTION:  USEFUL FOR MODIFYING PLANT CHARACTERISTICS
;  FILE REFERENCE: 2750-1601PUS2
;  CURRENT APPLICATION NUMBER: US/11/174,307B
;  CURRENT FILING DATE:  2005-06-30
;  PRIOR APPLICATION NUMBER: 60/583,671
;  PRIOR FILING DATE: 2004-06-30
;  PRIOR APPLICATION NUMBER: 60/583,781
;  PRIOR FILING DATE: 2004-06-30
;  PRIOR APPLICATION NUMBER: 60/583,651
;  PRIOR FILING DATE: 2004-06-30
;  NUMBER OF SEQ ID NOS: 5544
; SEQ ID NO 875
;   LENGTH: 2062
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ceres cDNA ID no. 12373375
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ceres CLONE  ID no. 259530
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Also known as  Ceres CDNA ID no. 24398502
US-11-174-307B-875

  Query Match             4.8%;  Score 60.2;  DB 9;  Length 2062;
  Best Local Similarity   45.4%;  Pred. No. 0.019;
  Matches  264;  Conservative    0;  Mismatches  308;  Indels    9;  Gaps    1;

Qy        648 GGAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCC 707
              || |  |||| ||| |||||   |||| |    | ||| |  ||  | | |   |||||
Db        888 GGCGGGCGTCATCGCCATCGATCTGGCCCGCGACATGCTCCAGGCCAGCGGCGCGGGCCT 947

Qy        708 CGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCC 767
               |    || |||| |||||| |  | || ||   |  |    || |    ||  |    |
Db        948 GGCCGTGGTGGTCAGCACGGAGGCCGTCTCCTTCACGTGGTACCCCGGGAAGCGGCGCTC 1007

Qy        768 CCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCT 827
              |  | || || |   || |  || |   | ||     || ||| ||||         |||
Db       1008 CATGCTCATCCCGAACGCCTTCTTCCGGGCCGGCTGCGCGGCGGTGCT---------GCT 1058

Qy        828 GGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCT 887
              | || ||     |   ||  ||  ||| |   |   |   || |    |||||||
Db       1059 GTCCAACCGGCGCCGCGACTTCCACCGCGCCAAGTACCAGCTGGAGCACGTGGTGCGCAC 1118

Qy        888 GGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCG 947
              |  | || |  |   || |  |  | |   |  | ||               |||  || ||
Db       1119 GCACAAGGGCGCGGACGACCGCGCCTTCCGCTCCGTGTACCAGGAGGAGGACGAGCAGCG 1178

Qy        948 GCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCT 1007
              | | | ||||        |  ||  ||| | || |  |     |     |||| ||  |
Db       1179 GATCAAGGGCCTGTCCATCAGCCGCGACCTGCTGGAGGTGGGCGGCCACGCGCTCAAGAC 1238

Qy       1008 GTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGA 1067
                | |    | |||||||     ||| ||| ||  |     || ||    | || |||
Db       1239 CAACATCACCACGCTGGGGCCCCTGGTGCTGCCCTTCTCCGAGCAGCTGCTCTTCTTCGC 1298

Qy       1068 ACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGC 1127
                         | ||||   |   |     || | || || | |   ||  |||   | |
```

```
Db       1299 GGGCGTGCTGTTCCGCCACCTGTTCCCGTCCAAGGCCTCCACGCCGGCGCCGCCGACCAC 1358

Qy       1128 CACCGCGGCTCCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGG 1187
              | | | ||   ||  |  | |  |  |   |  | ||| |  |  || |  | ||  |
Db       1359 CCCAGGGGACGCCTCGGCCGCCGCGCCCTACATCCCGGACTTCAAGCGCGCGTTCGAGCA 1418

Qy       1188 CGTACCCGGGCCCGCGCACACCCCCGCAGCGACGAACGAAC 1228
              | |   |  || ||||  || || ||  | |  | |  | |
Db       1419 CTTCTGCATGCACGCGGCCAGCCGCGACGTGCTGGAGCACC 1459


RESULT 11
US-11-056-355B-2961
; Sequence 2961, Application US/11056355B
; Publication No. US20060150283A1
; GENERAL INFORMATION:
;  APPLICANT: Brover, Vyacheslav
;  APPLICANT:  Alexandrov, Nickolai
;  TITLE OF INVENTION: Sequence Determined DNA Fragments and Corresponding
;  TITLE OF INVENTION:  Polypeptides Encoded Thereby
;  FILE REFERENCE: 2750-1590PUS2
;  CURRENT APPLICATION NUMBER: US/11/056,355B
;  CURRENT FILING DATE:  2005-02-14
;  PRIOR APPLICATION NUMBER: 60/544,190
;  PRIOR FILING DATE: 2004-02-13
;  NUMBER OF SEQ ID NOS: 119966
; SEQ ID NO 2961
;   LENGTH: 2062
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ceres Seq. ID no. 12373375
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ortholog of  Ceres SEQ ID NO 15175923
;   OTHER INFORMATION: as cited in SEQ ID NO 54037
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ortholog of  Ceres SEQ ID NO 14303366
;   OTHER INFORMATION: as cited in SEQ ID NO 54418
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ortholog of  Ceres SEQ ID NO 13503610
;   OTHER INFORMATION: as cited in SEQ ID NO 62416
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ortholog of  Ceres SEQ ID NO 13657610
;   OTHER INFORMATION: as cited in SEQ ID NO 66165
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ortholog of  Ceres SEQ ID NO 13598312
;   OTHER INFORMATION: as cited in SEQ ID NO 68448
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ortholog of  Ceres SEQ ID NO 13596100
;   OTHER INFORMATION: as cited in SEQ ID NO 68573
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(2062)
;   OTHER INFORMATION: Ortholog of  Ceres SEQ ID NO 15179123
;   OTHER INFORMATION: as cited in SEQ ID NO 0
US-11-056-355B-2961

  Query Match             4.8%;  Score 60.2;  DB 9;  Length 2062;
  Best Local Similarity   45.4%;  Pred. No. 0.019;
```

```
  Matches  264;  Conservative    0;  Mismatches  308;  Indels    9;  Gaps    1;

Qy        648 GGAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCC 707
              || |  |||| ||| |||||    |||| |    | ||| |  ||  | | |   |||||
Db        888 GGCGGGCGTCATCGCCATCGATCTGGCCCGCGACATGCTCCAGGCCAGCGGCGCGGGCCT 947

Qy        708 CGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCC 767
              |    || |||| |||||| |  | || ||   |  |   || |   ||  |    |
Db        948 GGCCGTGGTGGTCAGCACGGAGGCCGTCTCCTTCACGTGGTACCCCGGGAAGCGGCGCTC 1007

Qy        768 CCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCT 827
              |  | || || |   || |  || |    | ||      || ||| ||||          |||
Db       1008 CATGCTCATCCCGAACGCCTTCTTCCGGGCCGGCTGCGCGGCGGTGCT---------GCT 1058

Qy        828 GGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCT 887
              | || ||      |    ||  ||  ||| |   |   |    || |    |||||||
Db       1059 GTCCAACCGGCGCCGCGACTTCCACCGCGCCAAGTACCAGCTGGAGCACGTGGTGCGCAC 1118

Qy        888 GGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCG 947
              |  | || |   |   || |  | | |   | | |          |||  || ||
Db       1119 GCACAAGGGCGCGGACGACCGCGCCTTCCGCTCCGTGTACCAGGAGGAGGACGAGCAGCG 1178

Qy        948 GCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCT 1007
              | | | ||||       |  ||  ||| | || |   |      |    |||| ||  |
Db       1179 GATCAAGGGCCTGTCCATCAGCCGCGACCTGCTGGAGGTGGGCGGCCACGCGCTCAAGAC 1238

Qy       1008 GTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGA 1067
                | |    | |||||||     ||| ||| ||  |    || ||    | || |||
Db       1239 CAACATCACCACGCTGGGGCCCCTGGTGCTGCCCTTCTCCGAGCAGCTGCTCTTCTTCGC 1298

Qy       1068 ACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGC 1127
                     | ||||   |  |    || | || || | |  ||  ||| | |
Db       1299 GGGCGTGCTGTTCCGCCACCTGTTCCCGTCCAAGGCCTCCACGCCGGCGCCGCCGACCAC 1358

Qy       1128 CACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGG 1187
              | | | ||   ||  | | | |   | | ||| |  |  || |  | || |
Db       1359 CCCAGGGGACGCCTCGGCCGCCGCGCCCTACATCCCGGACTTCAAGCGCGCGTTCGAGCA 1418

Qy       1188 CGTACCCGGGCCCGCGCACACCCCCGCAGCGACGAACGAAC 1228
              | |   |  || ||||  || || ||  | |  | |  | |
Db       1419 CTTCTGCATGCACGCGGCCAGCCGCGACGTGCTGGAGCACC 1459
```

RESULT 12
US-10-449-902-4345
; Sequence 4345, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4345
; LENGTH: 938
; TYPE: DNA
; ORGANISM: Oryza sativa
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK104729
; DATABASE ENTRY DATE: 2002-08-28
US-10-449-902-4345

Query Match 4.8%; Score 59.8; DB 6; Length 938;
Best Local Similarity 45.6%; Pred. No. 0.021;

```
  Matches  211;  Conservative    0;  Mismatches  252;  Indels    0;  Gaps    0;

Qy        759 GATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGAT 818
              |||| ||| |||  || ||  |  | ||  |           |  || ||  |  |||||
Db        116 GATGATGCTCCCCGTCCTCCTCTCCTTCCTCCTGCTGCCATTCTCGTCCATGGCGCTGAC 175

Qy        819 CACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCT 878
              ||   |   || ||||||| ||  | || |   ||||  |  | | |   ||
Db        176 GCAGGATTTCTGCGTCGCCGACCTGACCTGCAGCGACACGCCGGCGGGGTACCCGTGCAA 235

Qy        879 GGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTA 938
              ||   | | |||    |   |||||||| | ||    ||      |  || | |
Db        236 GGCCAGCGTCGGCGCCGGCGACTTCGCCTACCACGGCCTCGCGGCGGCGGGCAACACCAG 295

Qy        939 CGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGC 998
              |   | |      |  | |||     | || || ||  | ||||    |  |  |  ||
Db        296 CAACCTCATCAAGGCCGCCGTGACGCCGGCCTTCGTCGGCCAGTTCCCCGGCGTGAACGG 355

Qy        999 GATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACAC 1058
              | |||   | |||  || || |||||   | || |||   | || ||  | ||| |||||
Db        356 GCTCGGTATCTCCGCGGCGAGGCTGGACATCGCCGTCGGCGGCGTCGTGCCGCTCCACAC 415

Qy       1059 CGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAA 1118
              | |  |      ||   | | |  ||  |    |||    |||| || |    |
Db        416 CCACCCGGCCGCCTCCGAGCTCCTGTTCGTCACCCAGGGCACCGTCGCCGCCGGGTTCAT 475

Qy       1119 GCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCT 1178
              ||  | || ||    |  |   | |   |  |  || ||  | |  | ||    |
Db        476 CACCTCCTCCTCCAACACCGTCTACACCAGGACGCTGTACGCCGGCGACATCATGGTGTT 535

Qy       1179 GCTCCCGGGCGTACCCGGGCCCGCGCACACCCCCGCAGCGACG 1221
              | || |||| | |       |  | ||| | |||     | ||
Db        536 CCCCCAGGGCCTCCTGCACTACCAGTACAACGCCGGCCAGTCG 578
```

RESULT 13
US-10-449-902-14734
; Sequence 14734, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14734
; LENGTH: 994
; TYPE: DNA
; ORGANISM: Oryza sativa
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK065284
; DATABASE ENTRY DATE: 2001-12-06
US-10-449-902-14734

Query Match 4.8%; Score 59.8; DB 6; Length 994;
Best Local Similarity 45.6%; Pred. No. 0.021;
Matches 211; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

```
Qy        759 GATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGAT 818
              |||| ||| |||  || ||  |  | ||  |           |  || ||  |  |||||
Db        105 GATGATGCTCCCCGTCCTCCTCTCCTTCCTCCTGCTGCCATTCTCGTCCATGGCGCTGAC 164

Qy        819 CACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCT 878
              ||   |   || ||||||| ||  | || |   ||||  |  | | |   ||
```

```
Db        165 GCAGGATTTCTGCGTCGCCGACCTGACCTGCAGCGACACGCCGGCGGGGTACCCGTGCAA 224

Qy        879 GGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTA 938

Db        225 GGCCAGCGTCGGCGCCGGCGACTTCGCCTACCACGGCCTCGCGGCGGCGGGCAACACCAG 284

Qy        939 CGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGC 998

Db        285 CAACCTCATCAAGGCCGCCGTGACGCCGGCCTTCGTCGGCCAGTTCCCCGGCGTGAACGG 344

Qy        999 GATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACAC 1058

Db        345 GCTCGGTATCTCCGCGGCGAGGCTGGACATCGCCGTCGGCGGCGTCGTGCCGCTCCACAC 404

Qy       1059 CGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAA 1118

Db        405 CCACCCGGCCGCCTCCGAGCTCCTGTTCGTCACCCAGGGCACCGTCGCCGCCGGGTTCAT 464

Qy       1119 GCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCT 1178

Db        465 CACCTCCTCCTCCAACACCGTCTACACCAGGACGCTGTACGCCGGCGACATCATGGTGTT 524

Qy       1179 GCTCCCGGGCGTACCCGGGCCCGCGCACACCCCCGCAGCGACG 1221

Db        525 CCCCCAGGGCCTCCTGCACTACCAGTACAACGCCGGCCAGTCG 567


RESULT 14
US-11-174-307B-4517
; Sequence 4517, Application US/11174307B
; Publication No. US20060143729A1
; GENERAL INFORMATION:
;  APPLICANT: ALEXANDROV, Nickolai
;  APPLICANT:  BROVER, Vyacheslav
;  TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND POLYPEPTIDES ENCODED THEREBY
;  TITLE OF INVENTION:  USEFUL FOR MODIFYING PLANT CHARACTERISTICS
;  FILE REFERENCE: 2750-1601PUS2
;  CURRENT APPLICATION NUMBER: US/11/174,307B
;  CURRENT FILING DATE:  2005-06-30
;  PRIOR APPLICATION NUMBER: 60/583,671
;  PRIOR FILING DATE: 2004-06-30
;  PRIOR APPLICATION NUMBER: 60/583,781
;  PRIOR FILING DATE: 2004-06-30
;  PRIOR APPLICATION NUMBER: 60/583,651
;  PRIOR FILING DATE: 2004-06-30
;  NUMBER OF SEQ ID NOS: 5544
; SEQ ID NO 4517
;   LENGTH: 1487
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1487)
;   OTHER INFORMATION: Ceres cDNA ID no. 21705438
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1487)
;   OTHER INFORMATION: Ceres CLONE  ID no. 1565945
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1487)
;   OTHER INFORMATION: Also known as  Ceres CDNA ID no. 25891794
US-11-174-307B-4517

  Query Match             4.8%;  Score 59.6;  DB 9;  Length 1487;
  Best Local Similarity   43.8%;  Pred. No. 0.024;
  Matches  306;  Conservative    0;  Mismatches  389;  Indels    3;  Gaps    1;

Qy        430 TCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATT 489

Db        246 TCGTACGCGGTGTTCGTGGCGCTGGCGCTGCTGGTGCCCGCCGCGGTGTCGCTGTCGCTG 305

Qy        490 CCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCC 549

Db        306 CGCGCCGACACGGCGCTGTCGCGGGCCTCCGCGTCCGCCATCGCCTTCAACCGCGTGGCG 365

Qy        550 CTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAG 609

Db        366 CAGGCCCCCGCCACGGGGCTCGCCGTCGTGTCCTTCGCCACGCTGGCCGCCTTCTTCCGC 425

Qy        610 GTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTG 669

Db        426 CGCTTCGGCGGCCTGCGGCAGCTGCTGTTCCTCGACGGCGCGCTCCGCGACGACACCGCC 485

Qy        670 ATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCG 729

Db        486 TTCGTGCGGCGCGGCTACGCGCGGGAGCTGGACCGCGCGTTCCGCACCCTGGCCGCGCTC 545

Qy        730 CTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCC 789

Db        546 CTGCTGCCCTCGCTCTGCGTGGAGGCCGCGCACAAGGCCGTCTTCTTCTTCTGCACCGTT 605

Qy        790 TGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGC 849

Db        606 AGGGTGGAGCCGCCCGCCGCCGTGCTCCTCCCCTTCGCCACGCCGCCGCTGGTGCTGCCG 665

Qy        850 CGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTC 909

Db        666 TGGCGCGCGGTGGCGCTGGTCGCCACGGTGGCGTCGTGGGTGTACCGCACGGGGGTGTTC 725

Qy        910 TACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGC 969

Db        726 CTGCTGGTGTGCGTGCTGTTCCGCCTCACCTGCGAGCTCCAGATCCTGCGCTTCGAGGGC 785

Qy        970 TACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTG 1029

Db        786 ATCTACCGCATGT---TCGACGTCGAGGCGCGCCCCACCGCCGCTGAGATCTTCGCGGAG 842

Qy       1030 GCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTC 1089

Db        843 CACCGCCGCATCCGCACGCAGCTGCTCGCCACCAGCCACCGGTACAGGGTGTTCATCCTC 902

Qy       1090 GGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGC 1127

Db        903 TGCTGCCTCGTCACCATCACCATCAGCCAGCTCGGCGC 940


RESULT 15
US-11-218-305-17428/c
; Sequence 17428, Application US/11218305
; Publication No. US20060141495A1
; GENERAL INFORMATION:
;  APPLICANT: MONSANTO TECHNOLOGY, LLC
;  APPLICANT:  McLaird, Paul L.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Wu, Kunsheng
;  TITLE OF INVENTION: Polymorphic Markers and Methods of Genotyping
;  TITLE OF INVENTION:  Corn.
;  FILE REFERENCE: 38-21 (53660)B
;  CURRENT APPLICATION NUMBER: US/11/218,305
;  CURRENT FILING DATE:  2005-09-01
;  PRIOR APPLICATION NUMBER: US 60/606,880
;  PRIOR FILING DATE: 2004-09-01
;  NUMBER OF SEQ ID NOS: 25043
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17428
;   LENGTH: 2219
;   TYPE: DNA
;   ORGANISM: Zea mays
US-11-218-305-17428

  Query Match             4.8%;  Score 59.2;  DB 9;  Length 2219;
  Best Local Similarity   45.3%;  Pred. No. 0.029;
  Matches  263;  Conservative    0;  Mismatches  308;  Indels    9;  Gaps    1;

Qy        649 GAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCC 708
```

```
Db       1274 GCGGGCGTCATCGCCATCGATCTGGCCCGCGACATGCTCCAGGCCAGCGGCGCGGGCCTG 1215

Qy        709 GGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCC 768

Db       1214 GCCGTGGTGGTCAGCACGGAGGCCGTCTCCTTCACGTGGTACCCCGGGAAGCGGCGCTCC 1155

Qy        769 CCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTG 828

Db       1154 ATGCTCATCCCGAACGCCTTCTTCCGGGCCGGCTGCGCGGCGGTGCT---------GCTG 1104

Qy        829 GCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTG 888

Db       1103 TCCAACCGGCGCCGCGACTTCCACCGCGCCAAGTACCAGCTGGAGCACGTGGTGCGCACG 1044

Qy        889 GGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGG 948

Db       1043 CACAAGGGCGCGGACGACCGCGCCTTCCGCTCCGTGTACCAGGAGGAGGACGAGCAGCGG 984

Qy        949 CTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTG 1008

Db        983 ATCAAGGGCCTGTCCATCAGCCGCGACCTGCTGGAGGTGGGCGGCCACGCGCTCAAGACC 924

Qy       1009 TCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAA 1068

Db        923 AACATCACCACGCTGGGGCCCCTGGTGCTGCCCTTCTCCGAGCAGCTGCTCTTCTTCGCG 864

Qy       1069 CAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCC 1128

Db        863 GGCGTGCTGTTCCGCCACCTGTTCCCGTCCAAGGCCTCCACGCCGGCGCCGCCGACCACC 804

Qy       1129 ACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGC 1188

Db        803 CCAGGGGACGCCTCGGCCGCCGCCCTACATCCCGGACTTCAAGCGCGCGTTCGAGCAC 744

Qy       1189 GTACCCGGGCCCGCGCACACCCCCGCAGCGACGAACGAAC 1228

Db        743 TTCTGCATGCACGCGGCCAGCCGCGACGTGCTGGAGCACC 704
```

Search completed: August 25, 2006, 22:45:55
Job time : 337 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2006, 21:35:03 ; Search time 907 Seconds
(without alignments)
9570.506 Million cell updates/sec

Title: US-10-680-860A-13
Perfect score: 1245
Sequence: 1 gtgctccagcgcgtcgatct.....aacccaccccgagaggatga 1245

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_8:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*
15: geneseqn2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1245	100.0	1245	12	ADM80046	Adm80046 Spiramyci
	2	1245	100.0	1245	12	ADN97562	Adn97562 S ambofac
	3	1245	100.0	11171	12	ADM80035	Adm80035 Spiramyci
	4	1245	100.0	11171	12	ADN97551	Adn97551 S ambofac
	5	794	63.8	2381	2	AAQ44449	Aaq44449 3-acylati
c	6	794	63.8	84428	12	ADM45913	Adm45913 Streptomy
	7	617.2	49.6	1810	2	AAQ14978	Aaq14978 AcyA gene
	8	356.8	28.7	1851	2	AAQ01760	Aaq01760 DNA seque
	9	356.4	28.6	84428	12	ADM45913	Adm45913 Streptomy
	10	351.2	28.2	1167	1	AAN92630	Aan92630 DNA fragm
	11	345	27.7	1646	3	AAZ35999	Aaz35999 S. kitasa
c	12	321.6	25.8	2729	1	AAN92629	Aan92629 DNA fragm
c	13	251.4	20.2	19016	8	AAD47222	Aad47222 Streptoco
	14	246.6	19.8	47988	10	ADE86070	Ade86070 Streptomy
	15	192.8	15.5	47981	4	AAF30757	Aaf30757 Micromono
	16	184	14.8	1137	8	AAL61192	Aal61192 Actinosyn
c	17	184	14.8	82746	8	AAL61224	Aal61224 Actinosyn
	18	86.6	7.0	1826	2	AAT63624	Aat63624 NruI-KpnI
	19	82.4	6.6	3957	5	AAA09686	Aaa09686 HSV-2 imm
	20	82.4	6.6	3957	10	ADG75121	Adg75121 Human her
	21	82.4	6.6	154746	6	AAD25519	Aad25519 Human her
c	22	82.4	6.6	154746	6	AAD25519	Aad25519 Human her
	23	82.4	6.6	154746	14	AEC07416	Aec07416 DNA of HS
c	24	82.4	6.6	154746	14	AEC07416	Aec07416 DNA of HS
c	25	79.4	6.4	114955	2	AAX53491	Aax53491 Human ade
c	26	78	6.3	39979	14	AED48907	Aed48907 S. hindus
	27	77.8	6.2	114955	2	AAX53491	Aax53491 Human ade
	28	70.6	5.7	1431	13	ADX62580	Adx62580 Plant ful
	29	69.8	5.6	65140	4	AAD17184	Aad17184 Streptomy
	30	69.8	5.6	125401	4	AAD17186	Aad17186 Streptomy
c	31	68.6	5.5	71138	12	ADO13830	Ado13830 dbv gene
	32	68.2	5.5	14055	8	AAL61170	Aal61170 Actinosyn
	33	66.6	5.3	1338	10	ABZ66715	Abz66715 Orthosomy
	34	66.6	5.3	15240	10	ABZ66809	Abz66809 Orthosomy
c	35	66.6	5.3	109519	5	AAS08693	Aas08693 Micromono
c	36	65.8	5.3	1381	12	ADO03179	Ado03179 Rice orth
c	37	65.8	5.3	1381	13	ADI43049	Adi43049 Plant tra
	38	65.8	5.3	50466	14	AED48540	Aed48540 S. fradia
c	39	65.2	5.2	90597	10	ADJ72363	Adj72363 Streptomy
c	40	65.2	5.2	90600	6	ABQ78872	Abq78872 S. roseos
	41	65	5.2	1650	8	ACA37821	Aca37821 Prokaryot
	42	65	5.2	24379	2	AAT93095	Aat93095 Streptomy
	43	65	5.2	24379	2	AAV25925	Aav25925 Streptomy
	44	64.8	5.2	3897	14	AED00556	Aed00556 HSV RS1 l
	45	64.8	5.2	3897	14	AED00555	Aed00555 HSV RS1 l

ALIGNMENTS

RESULT 1
ADM80046
ID ADM80046 standard; DNA; 1245 BP.
XX
AC ADM80046;
XX
DT 01-JUL-2004 (first entry)
XX
DE Spiramycin biosynthesis orf6*, SEQ ID 13.
XX
KW Antibacterial; Immunosuppressive; Antirheumatic; Antiarthritic;
KW Anthelmintic; Insecticide; spiramycin biosynthesis; macrolide;
KW spiramycin; antibiotic; immunosuppressant; autoimmune disease; gene; ds.
XX
OS Streptomyces ambofaciens.
XX
FH Key Location/Qualifiers
FT CDS 1. .1245
FT /*tag= a
FT /product= "orf6* protein"
XX
PN FR2845394-A1.
XX
PD 09-APR-2004.
XX
PF 08-OCT-2002; 2002FR-00012489.
XX
PR 08-OCT-2002; 2002FR-00012489.
XX
PA (AVET) AVENTIS PHARMA SA.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Blondelet RMH, Dominguez H, Darbon RE, Gerbaud C, Gondran A;
PI Karray F, Lacroix P, Oestreicher MBN, Pernodet JL, Tuphile K;
XX
DR WPI; 2004-307291/29.
DR P-PSDB; ADM80047.
XX
PT New polynucleotides encoding proteins involved in spiramycin
PT biosynthesis, useful for improving synthesis of macrolide antibiotics or
PT for generating new hybrid macrolides.

```
XX
PS   Claim 1; SEQ ID NO 13; 323pp; French.
XX
CC   The present invention relates to coding sequences (I) which encodes
CC   proteins involved in spiramycin (A) biosynthesis. (I) are used to improve
CC   production (optionally also purity) of macrolides (M), especially
CC   spiramycins, and to produce hybrid antibiotics. (M) are useful in human
CC   and veterinary medicine as antibiotics, but some are also
CC   immunosuppressants (e.g. FK506) useful for treating transplant rejection,
CC   rheumatoid arthritis and other autoimmune diseases; or insecticides and
CC   antihelmintics (e.g. avermectin). The present sequence encodes one such
CC   protein involved in spiramycin biosynthesis.
XX
SQ   Sequence 1245 BP; 141 A; 455 C; 410 G; 239 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1245;  DB 12;  Length 1245;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-189;
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60

Qy         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120

Qy        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180

Qy        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240

Qy        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300

Qy        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360

Qy        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420

Qy        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480

Qy        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540

Qy        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600

Qy        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660

Qy        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720

Qy        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780

Qy        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840

Qy        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900

Qy        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960

Qy        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020

Qy       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080

Qy       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140

Qy       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200

Qy       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245


RESULT 2
ADN97562
ID   ADN97562 standard; DNA; 1245 BP.
XX
AC   ADN97562;
XX
DT   15-JUL-2004  (first entry)
XX
DE   S ambofaciens spiramycin biosynthetic gene ORF6*.
XX
KW   ds; gene; antibacterial; immunosuppressive; antirheumatic; antiarthritic;
KW   anthelmintic; insecticide; biosynthesis; macrolide; antibiotic;
KW   spiramycin; veterinary medicine; immunosuppressant; transplant rejection;
KW   rheumatoid arthritis; autoimmune disease; insecticide; anthelmintic;
KW   avermectin.
XX
OS   Streptomyces ambofaciens.
XX
PN   WO2004033689-A2.
XX
PD   22-APR-2004.
XX
PF   08-OCT-2003; 2003WO-FR002962.
XX
PR   08-OCT-2002; 2002FR-00012489.
PR   27-FEB-2003; 2003FR-00002439.
PR   07-AUG-2003; 2003US-0493490P.
XX
PA   (AVET ) AVENTIS PHARMA SA.
PA   (CNRS ) CNRS.
XX
PI   Blondelet-Rouault M,  Dominguez H,  Darbon-Rongere E,  Gerbaud C;
PI   Gondran A,  Karray F,  Lacroix P,  Oestreicher-Mermet- Bouvier N;
PI   Pernodet J,  Tuphile K;
XX
DR   WPI; 2004-330455/30.
DR   P-PSDB; ADN97563.
XX
PT   New polynucleotides encoding proteins involved in spiramycin
PT   biosynthesis, useful for improving synthesis of macrolide antibiotics or
PT   for generating new hybrid macrolides.
XX
```

```
PS   Claim 1; SEQ ID NO 13; 480pp; French.
XX
CC   The invention relates to polynucleotides that encode polypeptides
CC   involved in biosynthesis of the macrolide antibiotic spiramycin (A), or
CC   equivalents within the degeneracy of the genetic code. The
CC   polynucleotides and polypeptides are used: (a) to improve production
CC   (optionally also purity) of macrolides (M), especially spiramycins; and
CC   (b) to produce hybrid antibiotics. The macrolides are useful in human and
CC   veterinary medicine as antibiotics, but some are also immunosuppressants
CC   (e.g. FK506) useful for treating transplant rejection, rheumatoid
CC   arthritis and other autoimmune diseases; or insecticides and
CC   anthelmintics (e.g. avermectin). This sequence corresponds to the open
CC   reading frames (ORF) encoding one of the polypeptides involved in
CC   biosynthetic pathway.
XX
SQ   Sequence 1245 BP; 141 A; 455 C; 410 G; 239 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1245;  DB 12;  Length 1245;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-189;
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60

Qy         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120

Qy        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180

Qy        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240

Qy        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300

Qy        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360

Qy        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420

Qy        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480

Qy        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540

Qy        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600

Qy        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660

Qy        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720

Qy        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780

Qy        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840

Qy        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900

Qy        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960

Qy        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020

Qy       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080

Qy       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140

Qy       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200

Qy       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245


RESULT 3
ADM80035
ID   ADM80035 standard; DNA; 11171 BP.
XX
AC   ADM80035;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Spiramycin biosynthesis related DNA, SEQ ID 2.
XX
KW   Antibacterial; Immunosuppressive; Antirheumatic; Antiarthritic;
KW   Anthelmintic; Insecticide; spiramycin biosynthesis; macrolide;
KW   spiramycin; antibiotic; immunosuppressant; autoimmune disease; gene; ds.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             710. .937
FT                   /*tag=  a
FT                   /product= "orf10* protein"
FT   CDS             1007. .1888
FT                   /*tag=  b
FT                   /product= "orf9* protein"
FT   CDS             1925. .2755
FT                   /*tag=  c
FT                   /product= "orf8* protein"
FT   CDS             complement(2817. .3665)
FT                   /*tag=  d
FT                   /product= "orf7*c protein"
FT   CDS             4061. .5305
FT                   /*tag=  e
FT                   /product= "orf6* protein"
FT   CDS             5302. .5976
FT                   /*tag=  f
FT                   /product= "orf5* protein"
FT   CDS             complement(6100. .7449)
FT                   /*tag=  g
FT                   /product= "orf4*c protein"
FT   CDS             complement(7476. .8741)
FT                   /*tag=  h
```

```
FT                   /product= "orf3*c protein"
FT   CDS             complement(8781. .10052)
FT                   /*tag=  i
FT                   /product= "orf2*c protein"
FT   CDS             complement(10172. .10882)
FT                   /*tag=  j
FT                   /product= "orf1*c protein"
XX
PN   FR2845394-A1.
XX
PD   09-APR-2004.
XX
PF   08-OCT-2002; 2002FR-00012489.
XX
PR   08-OCT-2002; 2002FR-00012489.
XX
PA   (AVET ) AVENTIS PHARMA SA.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Blondelet RMH,  Dominguez H,  Darbon RE,  Gerbaud C,  Gondran A;
PI   Karray F,  Lacroix P,  Oestreicher MBN,  Pernodet JL,  Tuphile K;
XX
DR   WPI; 2004-307291/29.
DR   P-PSDB; ADM80037, ADM80039, ADM80041, ADM80043, ADM80045, ADM80047,
DR   ADM80049, ADM80051, ADM80053, ADM80055.
XX
PT   New polynucleotides encoding proteins involved in spiramycin
PT   biosynthesis, useful for improving synthesis of macrolide antibiotics or
PT   for generating new hybrid macrolides.
XX
PS   Example 3; SEQ ID NO 2; 323pp; French.
XX
CC   The present invention relates to coding sequences (I) which encodes
CC   proteins involved in spiramycin (A) biosynthesis. (I) are used to improve
CC   production (optionally also purity) of macrolides (M), especially
CC   spiramycins, and to produce hybrid antibiotics. (M) are useful in human
CC   and veterinary medicine as antibiotics, but some are also
CC   immunosuppressants (e.g. FK506) useful for treating transplant rejection,
CC   rheumatoid arthritis and other autoimmune diseases; or insecticides and
CC   antihelmintics (e.g. avermectin). The present sequence encodes a number
CC   of proteins involved in spiramycin biosynthesis.
XX
SQ   Sequence 11171 BP; 1431 A; 3878 C; 4182 G; 1680 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1245;  DB 12;  Length 11171;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-189;
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 4120

Qy         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 4180

Qy        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 4240

Qy        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 4300

Qy        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 4360

Qy        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 4420

Qy        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 4480

Qy        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 4540

Qy        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 4600

Qy        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 4660

Qy        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 4720

Qy        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 4780

Qy        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 4840

Qy        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 4900

Qy        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 4960

Qy        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 5020

Qy        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 5080

Qy       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5081 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 5140

Qy       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5141 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 5200

Qy       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5201 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 5260

Qy       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       5261 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 5305


RESULT 4
ADN97551
ID   ADN97551 standard; DNA; 11171 BP.
XX
AC   ADN97551;
XX
DT   15-JUL-2004  (first entry)
XX
DE   S ambofaciens spiramycin biosynthetic enzyme genomic region #2.
XX
KW   ds; gene; antibacterial; immunosuppressive; antirheumatic; antiarthritic;
KW   anthelmintic; insecticide; biosynthesis; macrolide; antibiotic;
KW   spiramycin; veterinary medicine; immunosuppressant; transplant rejection;
```

```
KW   rheumatoid arthritis; autoimmune disease; insecticide; anthelmintic;
KW   avermectin.
XX
OS   Streptomyces ambofaciens.
XX
PN   WO2004033689-A2.
XX
PD   22-APR-2004.
XX
PF   08-OCT-2003; 2003WO-FR002962.
XX
PR   08-OCT-2002; 2002FR-00012489.
PR   27-FEB-2003; 2003FR-00002439.
PR   07-AUG-2003; 2003US-0493490P.
XX
PA   (AVET ) AVENTIS PHARMA SA.
PA   (CNRS ) CNRS.
XX
PI   Blondelet-Rouault M,  Dominguez H,  Darbon-Rongere E,  Gerbaud C;
PI   Gondran A,  Karray F,  Lacroix P,  Oestreicher-Mermet- Bouvier N;
PI   Pernodet J,  Tuphile K;
XX
DR   WPI; 2004-330455/30.
XX
PT   New polynucleotides encoding proteins involved in spiramycin
PT   biosynthesis, useful for improving synthesis of macrolide antibiotics or
PT   for generating new hybrid macrolides.
XX
PS   Example 4; SEQ ID NO 2; 480pp; French.
XX
CC   The invention relates to polynucleotides that encode polypeptides
CC   involved in biosynthesis of the macrolide antibiotic spiramycin (A), or
CC   equivalents within the degeneracy of the genetic code. The
CC   polynucleotides and polypeptides are used: (a) to improve production
CC   (optionally also purity) of macrolides (M), especially spiramycins; and
CC   (b) to produce hybrid antibiotics. The macrolides are useful in human and
CC   veterinary medicine as antibiotics, but some are also immunosuppressants
CC   (e.g. FK506) useful for treating transplant rejection, rheumatoid
CC   arthritis and other autoimmune diseases; or insecticides and
CC   anthelmintics (e.g. avermectin). This sequence corresponds to a genomic
CC   fragment containing the open reading frames (ORF) of the genes encoding
CC   the polypeptides involved in biosynthetic pathway. This fragment overlaps
CC   the fragment shown in ADN97689 to cover the new ORF of the invention.
XX
SQ   Sequence 11171 BP; 1431 A; 3878 C; 4182 G; 1680 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1245;  DB 12;  Length 11171;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-189;
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 4120

Qy         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 4180

Qy        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 4240

Qy        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 4300

Qy        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 4360

Qy        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 4420

Qy        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 4480

Qy        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 4540

Qy        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 4600

Qy        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 4660

Qy        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 4720

Qy        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 4780

Qy        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 4840

Qy        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 4900

Qy        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 4960

Qy        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 5020

Qy        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 5080

Qy       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5081 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 5140

Qy       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5141 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 5200

Qy       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5201 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 5260

Qy       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       5261 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 5305


RESULT 5
AAQ44449
ID   AAQ44449 standard; DNA; 2381 BP.
XX
AC   AAQ44449;
XX
DT   11-OCT-1994  (first entry)
XX
DE   3-acylating enzyme coding sequence.
XX
KW   Enzyme; acylation; oxygen; position 3; macrolide; antibiotic;
```

```
KW   plasmid pIJ680; transformation; culture; acylating enzyme; ss.
XX
OS   Streptomyces mycarofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             126. .1289
FT                   /*tag=  a
FT                   /product= "3-acylating enzyme"
XX
PN   JP06038750-A.
XX
PD   15-FEB-1994.
XX
PF   09-AUG-1991;   91JP-00224814.
XX
PR   09-AUG-1991;   91JP-00224814.
XX
PA   (MEIJ ) MEIJI SEIKA KAISHA.
XX
DR   WPI; 1994-094826/12.
DR   P-PSDB; AAR49828.
XX
PT   3-acylating enzyme for producing macrolide antibiotics - has specified
PT   amino acid sequence of 387 residues.
XX
PS   Claim 3; Page 8-10; 13pp; Japanese.
XX
CC   This sequence encodes an enzyme for acylating oxygen at position 3 of a
CC   macrolide antibiotic. This enzyme can be used for the creation of new
CC   macrolide antibiotics. Streptomyces mycarofaciens SF-837 was cultured in
CC   a TSB medium 30 deg. C for 24 hrs. It was subcultured in a SGGP medium at
CC   30 deg. C for 24 hrs. The culture was centrifuged and the microbe body
CC   was extracted and the total DNA was dialysed to give the donor DNA. It
CC   was purified and ligated with plasmid pIJ680 to prepare recombinant DNA.
CC   S. lividans was transformed by the resulting plasmid. This transformant
CC   can be cultured to produce large amounts of the acylating enzyme
XX
SQ   Sequence 2381 BP; 341 A; 813 C; 783 G; 444 T; 0 U; 0 Other;

  Query Match             63.8%;  Score 794;  DB 2;  Length 2381;
  Best Local Similarity   80.9%;  Pred. No. 2.5e-117;
  Matches  937;  Conservative    0;  Mismatches  220;  Indels    1;  Gaps    1;

Qy         12 CGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTG 71
              ||||   ||| |||| |||||||||| |||||||  |||||   |  ||||||||||||
Db        140 CGTCCGCCTGCCGTCGCTCACCGGCCTTCGCTGGTTCGCGGCCTTAGCGGTATTCGCCTG 199

Qy         72 TCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACAT 131
              ||||||  |||||||||    |||||||| |||||||||| || | ||| ||||| |||||
Db        200 TCACATTGCCCAGCAGCAATTCTTCGCTGACCAGCAGGTCGGAACCGCGCTGCTACACAT 259

Qy        132 CACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGC 191
              |||| |||||||||| || |||||||| ||||||| | || || || || || || || ||
Db        260 CACCACGCTCGGTTCGATCGCGGTCTCGGTCTTCTTCCTGCTCAGCGGTTTTGTTCTGGC 319

Qy        192 CTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGAT 251
              |||||||||| ||| ||||||| ||  ||||||||||||||||||  | || |||||
Db        320 GTGGTCGGCCCGTGACAAGGACTCTGTCACGACTTTCTGGCGGCGCCGATTTGCCAAGAT 379

Qy        252 CTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGA 311
              ||| |||||||| |||| || |||  ||| |||| |||||||||||||||  ||||
Db        380 CTACCCGCTGCACCTCGTAACTTTCCTCATAGCGGGCGTCATCATTTTCTCCCTCGCGGA 439

Qy        312 GCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCA 371
              ||||   || ||||| |||||||||||||||| ||||| ||| || ||||||| |||||
Db        440 GCCGACTCTGCCCGGTGGTTCCGTATGGGACGGTCTGGTTCCCGATCTTCTGCTGGTGCA 499

Qy        372 GTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTC 431
              ||||||||||||||  | ||  | |  ||| | |||||||| || || ||||||||||
Db        500 GTCCTGGCTTCCCGAACCCACGATTATCGCCGGGTTCAACACCCCGAGTTGGTCGCTCTC 559

Qy        432 CTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCC 491
              ||||||  |||| ||||||||| |||||||||||||||| |||||||||| | ||||||
Db        560 CTGTGAATTCGCTTTCTATCTGACGTTCCCGCTGTGGTATCGGCTGGTGCGGAAGATTCC 619

Qy        492 CGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCT 551
              | |||||||||||||||||| ||||| || || |||| |||||||  ||||| ||  |
Db        620 GGTACGGCGGCTGTGGTGGTGTGCCGCCGGTATTGCCGCGGCCGTGATTTGTGTACCGTT 679

Qy        552 GCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGT 611
              || || ||| | |||||||||||  |||  |||||||||||||| |||||||||| |
Db        680 CGTGACGAGCCAGTTCCCGGCGAGCGCCGAGACGGCCCCCGGGATGCCGCTCAACGAGCT 739

Qy        612 CTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGAT 671
              |||||||| | |||||| |||||||||| ||||||||||||||||||||||||||||||
Db        740 GTGGTTCGCCTGCTGGCTGCCGCCGGTGCGGATGCTGGAGTTCGTCCTCGGCATCGTGAT 799

Qy        672 GGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCT 731
              |||||||||||||||| ||||  | |||  |||||||||   || |  || | ||  |||
Db        800 GGCGCTGATCCTGCGCACGGGCGTGTGGCGGGGCCCCGGAGTGGTGTCCTCCGCGCTGCT 859

Qy        732 CCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTG 791
              |||||||||   |||||| |||||||| |||||||||||| |||||| ||| ||| ||
Db        860 GCTCGCCGCGGCCTACGGCGTCACCCAGGTGGTGCCCCCGATGTTCACCATCGCCGCGTG 919

Qy        792 CTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCG 851
              |||| |||| || ||||| |||||||||| ||  ||||| | ||||| ||||| ||||
Db        920 CTCCATCGTGCCCGCCGCCCTGCTGATCACCGCTTTGGCCAATGCCGATGTGCAGGGCCT 979

Qy        852 GCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTA 911
              |||||| ||  |||| ||||||  |||||||||||||| ||||||||||||||||||||
Db        980 GCGCACCGGTTTGCGCTCGGCGGTACTGGTGCGGCTGGGCGAGTGGTCCTTCGCCTTCTA 1039

Qy        912 CCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTA 971
              ||||| |||||| || |||||||||||||| ||||||||||||||||||||| ||||||||
Db       1040 TCTGGTTCACTTCATGGTCATCCGCTACGGGCACCGGCTGATGGGCGGCGAACTGGGCTA 1099

Qy        972 CGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGC 1031
              || ||| |||||||||||| ||  ||||   |||||| |||||| |||||||| ||| ||
Db       1100 CGCGCGCCAGTGGAGCACCGCGAGCGCGGGAGCGCTGGCCCTGGCGATGCTGGCGGTCGC 1159

Qy       1032 GGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGG 1091
              | || ||||||| || |||||||||||||||||  ||  | |||||||||| ||| ||||
Db       1160 GATCGTGGCCGGTGGGCTGCTGCACACCGTCGTGGAGAACCCCTGCATGCGTCTGCTCGG 1219

Qy       1092 CAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG-CGGCTCCCCGGAACTCAC 1150
              | ||||| |  | | | ||     |       |  ||| |||| ||   | ||| |||||||
Db       1220 CCGCCGCCGCCCGGTCGCCACCGCACCAGACCCCGCAACCGACGAGGCTCCGAAACTCAC 1279

Qy       1151 CCGCGGCCGACGCGGCCG 1168
              ||||| | ||||   |||
Db       1280 CCGCGCCTGACGTCACCG 1297


RESULT 6
ADM45913/c
ID   ADM45913 standard; DNA; 84428 BP.
XX
AC   ADM45913;
XX
DT   03-JUN-2004  (first entry)
XX
DE   Streptomyces mycarofaciens midecamycin polyketide synthetase DNA.
XX
KW   macrolide; midecamycin biosynthesis; polyketide synthetase; PKS;
KW   antimicrobial; ds; gene.
XX
OS   Streptomyces mycarofaciens.
XX
PN   JP2004049100-A.
XX
PD   19-FEB-2004.
XX
PF   19-JUL-2002; 2002JP-00210516.
XX
PR   19-JUL-2002; 2002JP-00210516.
```

```
XX
PA   (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX
DR   WPI; 2004-174115/17.
DR   P-PSDB; ADM45914, ADM45915, ADM45916, ADM45917, ADM45918, ADM45919,
DR   ADM45920, ADM45921, ADM45922, ADM45923, ADM45924, ADM45925, ADM45926,
DR   ADM45927, ADM45928, ADM45929, ADM45930, ADM45931, ADM45932, ADM45933,
DR   ADM45934, ADM45935, ADM45936, ADM45937, ADM45938, ADM45939, ADM45940,
DR   ADM45941, ADM45942, ADM45943, ADM45944, ADM45945, ADM45946, ADM45947,
DR   ADM45948, ADM45949, ADM45950, ADM45951, ADM45952, ADM45953, ADM45954,
DR   ADM45955.
XX
PT   New polynucleotides encoding proteins important in the biosynthesis of
PT   midecamycin.
XX
PS   Claim 2; SEQ ID NO 1; 264pp; Japanese.
XX
CC   The invention relates to a novel polynucleotide, specifically a
CC   biosynthesis gene, comprising a nucleotide sequence which encodes a
CC   protein, where the gene is concerned with the biosynthesis of
CC   midecamycin. The polynucleotide of the invention demonstrates
CC   antimicrobial activity and may be useful for the biosynthesis of
CC   midecamycin, as well as for manufacturing macrolide compounds other than
CC   midecamycin. The current sequence is that of the Streptomyces
CC   mycarofaciens midecamycin polyketide synthetase (PKS) DNA of the
CC   invention.
XX
SQ   Sequence 84428 BP; 11456 A; 29098 C; 30275 G; 13599 T; 0 U; 0 Other;

  Query Match             63.8%;  Score 794;  DB 12;  Length 84428;
  Best Local Similarity   80.9%;  Pred. No. 1.7e-117;
  Matches  937;  Conservative    0;  Mismatches  220;  Indels    1;  Gaps    1;

Qy         12 CGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTG 71
Db      77751 CGTCCGCCTGCCGTCGCTCACCGGCCTTCGCTGGTTCGCGGCCTTAGCGGTATTCGCCTG 77692

Qy         72 TCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACAT 131
Db      77691 TCACATTGCCCAGCAGCAATTCTTCGCTGACCAGCAGGTCGGAACCGCGCTGCTACACAT 77632

Qy        132 CACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGC 191
Db      77631 CACCACGCTCGGTTCGATCGCGGTCTCGGTCTTCTTCCTGCTCAGCGGTTTTGTTCTGGC 77572

Qy        192 CTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGAT 251
Db      77571 GTGGTCGGCCCGTGACAAGGACTCTGTCACGACTTTCTGGCGGCGCCGATTTGCCAAGAT 77512

Qy        252 CTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGA 311
Db      77511 CTACCCGCTGCACCTCGTAACTTTCCTCATAGCGGGCGTCATCATTTTCTCCCTCGCGGA 77452

Qy        312 GCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCA 371
Db      77451 GCCGACTCTGCCCGGTGGTTCCGTATGGGACGGTCTGGTTCCCGATCTTCTGCTGGTGCA 77392

Qy        372 GTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTC 431
Db      77391 GTCCTGGCTTCCCGAACCCACGATTATCGCCGGGTTCAACACCCCGAGTTGGTCGCTCTC 77332

Qy        432 CTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCC 491
Db      77331 CTGTGAATTCGCTTTCTATCTGACGTTCCCGCTGTGGTATCGGCTGGTGCGGAAGATTCC 77272

Qy        492 CGCACGGCGGCTGTGGTGGTGCCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCT 551
Db      77271 GGTACGGCGGCTGTGGTGGTGTGCCGCCGGTATTGCCGCGGCCGTGATTTGTGTACCGTT 77212

Qy        552 GCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGT 611
Db      77211 CGTGACGAGCCAGTTCCCGGCGAGCGCCGAGACGGCCCCCGGGATGCCGCTCAACGAGCT 77152

Qy        612 CTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGAT 671
Db      77151 GTGGTTCGCCTGCTGGCTGCCGCCGGTGCGGATGCTGGAGTTCGTCCTCGGCATCGTGAT 77092

Qy        672 GGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCT 731
Db      77091 GGCGCTGATCCTGCGCACGGGCGTGTGGCGGGGCCCCGGAGTGGTGTCCTCCGCGCTGCT 77032

Qy        732 CCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTG 791
Db      77031 GCTCGCCGCGGCCTACGGCGTCACCCAGGTGGTGCCCCCGATGTTCACCATCGCCGCGTG 76972

Qy        792 CTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCG 851
Db      76971 CTCCATCGTGCCCGCCGCCCTGCTGATCACCGCTTTGGCCAATGCCGATGTGCAGGGCCT 76912

Qy        852 GCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTA 911
Db      76911 GCGCACCGGTTTGCGCTCGGCGGTACTGGTGCGGCTGGGCGAGTGGTCCTTCGCCTTCTA 76852

Qy        912 CCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTA 971
Db      76851 TCTGGTTCACTTCATGGTCATCCGCTACGGGCACCGGCTGATGGGCGGCGAACTGGGCTA 76792

Qy        972 CGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGC 1031
Db      76791 CGCGCGCCAGTGGAGCACCGCGAGCGCGGGAGCGCTGGCCCTGGCGATGCTGGCGGTCGC 76732

Qy       1032 GGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGG 1091
Db      76731 GATCGTGGCCGGTGGGCTGCTGCACACCGTCGTGGAGAACCCCTGCATGCGTCTGCTCGG 76672

Qy       1092 CAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG-CGGCTCCCCGGAACTCAC 1150
Db      76671 CCGCCGCCGCCCGGTCGCCACCGCACCAGACCCCGCAACCGACGAGGCTCCGAAACTCAC 76612

Qy       1151 CCGCGGCCGACGCGGCCG 1168
Db      76611 CCGCGCCTGACGTCACCG 76594


RESULT 7
AAQ14978
ID   AAQ14978 standard; DNA; 1810 BP.
XX
AC   AAQ14978;
XX
DT   24-OCT-2003  (revised)
DT   25-FEB-1992  (first entry)
XX
DE   AcyA gene.
XX
KW   Macrolide antibiotics; acylation; Streptomyces; tylosin; spiramycin;
KW   leucomycin;. nidamycin; ss.
XX
OS   Streptomyces thermotolerans; ATCC 11416.
XX
FH   Key             Location/Qualifiers
FT   -35_signal      120. .125
FT                   /*tag=  a
FT   -10_signal      143. .148
FT                   /*tag=  b
FT   RBS             194. .198
FT                   /*tag=  c
FT   RBS             201. .205
FT                   /*tag=  d
FT   CDS             209. .1375
FT                   /*tag=  e
XX
PN   EP459525-A.
XX
PD   04-DEC-1991.
XX
```

```
PF   03-JUN-1991;   91EP-00109033.
XX
PR   01-JUN-1990;   90JP-00141625.
PR   22-FEB-1991;   91JP-00048753.
XX
PA   (SAOC ) MERCIAN CORP.
XX
PI   Arisawa A,  Kawamura N,  Kojima I,  Tone H,  Okamoto R,  Okamura K;
XX
DR   WPI; 1991-355809/49.
DR   P-PSDB; AAR15428.
XX
PT   Gene encoding 3-acylation enzyme for macrolide antibiotics - used to
PT   transform Streptomyces sp. to have acylation activity, for use in
PT   medicine.
XX
PS   Claim 7; Fig 7; 45pp; English.
XX
CC   The sequence was obtd. from plasmid p53A, a subclone of a DNA library
CC   prepd. from genomic DNA of S. thermotolerans. The gene encodes a 3-
CC   acylation enzyme involved in the synthesis of macrolide antibiotics.
CC   Microorganisms, ep. Streptomyces, transformed with the gene can be used
CC   for the commercial prodn. of macrolide antibiotics, e.g. 3-acylated
CC   tylosin. (Updated on 24-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1810 BP; 237 A; 633 C; 611 G; 329 T; 0 U; 0 Other;

  Query Match             49.6%;  Score 617.2;  DB 2;  Length 1810;
  Best Local Similarity   70.8%;  Pred. No. 2.9e-89;
  Matches  820;  Conservative    0;  Mismatches  338;  Indels    0;  Gaps    0;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
Db        230 CTACCTTCACTGACCGGGCTGCGCTGGTTCGCGGCGCTTTCCGTATTCGTCTGCCATATC 289

Qy         79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCCCG 138
Db        290 GCCCAGCAGGGCATTTTCGCCGACCCGGACGTGGCGTCCGCCCTGGGGCACCTCACGTCG 349

Qy        139 CTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCG 198
Db        350 CTCGGCTCGCTCGCGGTCTCCCTCTTCTTCGTGCTGAGCGGCTATGTGCTGACGTGGTCG 409

Qy        199 GCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCG 258
Db        410 GCCCGGGACGGCGACTCCGTCAGAAGCTTCTGGCAGCGTAGGTTCGCCAAGATCTACCCG 469

Qy        259 CTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCGGTA 318
Db        470 CTGCATTTCGTCACCTTCTGCATTGCGGGCGTCATCATCTTCAGTCCTTCCGAGCCGGTG 529

Qy        319 CTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCCTGG 378
Db        530 CTGCCGGGCGGTTCGACGTGGGACGGCATGGTGCCCAACCTGCTGCTGGTGCATTCATGG 589

Qy        379 CTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGTGAG 438
Db        590 CTGCCGGACGCGTACATCGTCTCGGGATTCAACACACCGAGCTGGTCGCTTTCCTGTGAA 649

Qy        439 ATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGG 498
Db        650 ATGGCCTTCTATCTCACGTTTCCGTTGTGGTACCGGCTGCTGCTTCGGATACGGGTGAGC 709

Qy        499 CGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCG 558
Db        710 CGGTTGTGGTGGTACGCGGCCGCACTGGCGCTGGCCGTGGTGTGCATGCCGTTCGTGGCC 769

Qy        559 GGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTC 618
Db        770 CGGCTGCTGCCGGACAGCGCGGAGGTCGTTCCCGGGATGCCGCTGCGGGACATGTGGTTC 829

Qy        619 GCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTG 678
Db        830 ACGTACTGGTTCCCGCCCGTGCGGATGCTCGAGTTCCTCCTCGGCATCGTGCTGGCGCTG 889

Qy        679 ATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCC 738
Db        890 ATCCCGCGCCAGGGGGCGTGGCGGGGGCCCGGAACGGGCACGGCCGCGCTGCTGCTCGGC 949

Qy        739 GCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTC 798
Db        950 GGCGCGTTCGCGCTCAACCAGGTGGTGCCGCCGATGTTCACCCTCACCGCCACCACCGTC 1009

Qy        799 GTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACG 858
Db       1010 CTCCCCATCGCCCTGCTGATCGCCGCCGCGGCGGACGGCGATCTGCGCGGGCGCCGTACC 1069

Qy        859 GGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTC 918
Db       1070 GGACTGGGCGCGGCCGTCCTGGTCAGGCTGGGCGAGTGGTCGTACGCCTTCTACCTGATC 1129

Qy        919 CACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGG 978
Db       1130 CACTTCCTGATCATTCGCTACGGGCACCGGCTGCTGGGCGGCGACCAGGGATACGCCCGG 1189

Qy        979 CAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTG 1038
Db       1190 CAGTGGGACACCCTCGCGGCGCTCGGCATCACAGCGGCGGTACTGGGGGTCACGATCGCC 1249

Qy       1039 GCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGC 1098
Db       1250 GCGAGCGCGGTCCTGCACTTCTTCGTCGAGCGGCCCTGTATGACCCTCCTGCGCGGCCGC 1309

Qy       1099 AGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCC 1158
Db       1310 CGCCCTCCGCAGGGGCCGGCTCCCGACTCCGGGGGCCGCCCGCACCGGGCTCCGCTGGAA 1369

Qy       1159 GACGCGGCCGGCGTGCCC 1176
Db       1370 AGGGCATGACGCGTGGCC 1387


RESULT 8
AAQ01760
ID   AAQ01760 standard; DNA; 1851 BP.
XX
AC   AAQ01760;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   02-AUG-1990  (first entry)
XX
DE   DNA sequence of a novel gene, carbomycin biosynthetic gene (carE), that
DE   encodes a 4"-O-isovaleryl acylase enzyme.
XX
KW   4"-O-isovaleryl acylase enzyme; carbomycin biosynthetic gene (carE).
XX
OS   Streptomyces thermotolerans.
XX
FH   Key             Location/Qualifiers
FT   CDS             581. .1747
FT                   /*tag=  a
FT   misc_feature    581. .1744
FT                   /*tag=  b
FT                   /note= "claimed in claim 7"
XX
PN   EP354641-A.
XX
PD   14-FEB-1990.
XX
PF   10-MAY-1989;   89EP-00304716.
XX
PR   13-MAY-1988;   88US-00194672.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
```

PI Epp JK, Schoner BE;
XX
DR WPI; 1990-046338/07.
DR P-PSDB; AAR05225.
XX
PT Car E gene encoding 4"-O-isovaleryl:acylase - used to transform cells to
PT increase prodn. of carbomycin or other antibiotics or to produce new
PT antibiotics.
XX
PS Disclosure; Page ?; 28pp; English.
XX
CC It was isolated from a carbomycin-producing strain of S. thermotolerans
CC ATCC 11416. It may also be isloated from S. griseofuscus. It is useful
CC for increasing 4"-O-isovaleryl-acylase (IVA) enzyme activity in an
CC organism. This can be used for increasing the yield of carbomycin and
CC other antibiotics in fermentations and also for producing new antibiotics
CC and antibiotic derivs. (Updated on 25-MAR-2003 to correct PA field.)
CC (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ Sequence 1851 BP; 263 A; 653 C; 576 G; 359 T; 0 U; 0 Other;

Query Match 28.7%; Score 356.8; DB 2; Length 1851;
Best Local Similarity 58.0%; Pred. No. 6.3e-48;
Matches 711; Conservative 0; Mismatches 502; Indels 13; Gaps 4;

```
Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61
              ||| || ||      |||| |  | ||| |||  | || |       | || ||   ||
Db        582 TGCCCCTGCCGAAACATCTTCCCGCGCTCGGCGGGATGCGTTTCATCTCCGCTCTACTGG 641

Qy         62 TATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCAC 121
              ||||| |||  || ||  |    |||   || |||    ||    || |      ||| |
Db        642 TATTCACCTCCCATATATCGACACAGCCGTTATTCAAGAACACCGAGATCAATTCCGCGC 701

Qy        122 TGC---TGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTG 178
              |||   | |   | | || ||| ||  |  || |||||||| |||||||||| || || |
Db        702 TGCAGTTCCCGCTGAACCGGCTGGGCCCGCTGACGGTCTCGTTCTTCTTCATGCTCAGCG 761

Qy        179 GATTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCC 238
              | |||||||| ||||| |||  |    |||  | |||  |  ||  |||||||||| |
Db        762 GTTTCGTCCTCACCTGGGCGGGTCTGCCCGACAAGTCCAAGGTGAACTTCTGGCGGCGGC 821

Qy        239 GCATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTT 298
              |||  |        ||  ||||||| ||  |   | |     |  ||   || ||| |
Db        822 GCACGGTCCGCGCGTACTCGCTGCACCTGCCCGTGCTGCTGGTGACGCTGCTGATCGTGC 881

Qy        299 TCTCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATG 358
              |  ||||    |||||      |    ||| | || || |||||||| ||| |  | ||
Db        882 TGGCCCTCAACGAGCCCAACAT---GGGCCGATCGGTGTGGGACGGACTGCTCACGAACC 938

Qy        359 TTCTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCA 418
              | |||||  | ||| | ||| | ||||||      |      | |||  | ||| ||
Db        939 TGCTGCTGATCCAGGCATGGTTCCCCGACCACCACGAGTACGGCAGCATGAACCCGGTGG 998

Qy        419 GCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGG 478
              ||||||||||||||| ||| |   ||||||       |||||||  |    | |    |
Db        999 CGTGGTCGCTCTCCTGCGAGCTGTTCTTCTACGCCATGTTCCCGTTCCTCTTCGCCTTCT 1058

Qy        479 TGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGA 538
              |   || | | |    |      ||||| ||| |||| ||||| | | |  |||||||||
Db       1059 TCACCAAGGTCCGTACGGACCGGCTCTGGCGGTGGGCCGCCGCGGTGTCCGTGGCCGCCG 1118

Qy        539 CGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAG------CGAGGAGGTGGCCCCCG 592
              |   | ||||||  | ||    ||||| ||||| ||      |  |  |  || ||| |
Db       1119 TCTCCATCCCCCTGGTCGCACTGCTGCTGCCGGCCAGCCCGCCCCTGCCGTGGGACCCGG 1178

Qy        593 GGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGT 652
              ||| |||       |   ||||||   |||  | | ||||||||||  |||||||||
Db       1179 ACATGCCGCAGCTGCGGTGGTGGTTCATCTACATGTTCCCGCCGGTGCGGCTGCTGGAGT 1238

Qy        653 TCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712
              |||| ||||| ||  | ||||| | |||| ||  |  |||   ||||| ||||||  |||
Db       1239 TCGTGCTCGGGATGCTCATGGCCCAGATCGTGATCCGGGGACGCTGGAGGGGCCCGCGTC 1298

Qy        713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772
              |   || ||||   |||||  || | |||   | ||   | ||       ||||||
Db       1299 CCCTGGCCTGCGTCGCGCTGTTCTCAGCGGTGTTCGCGGTGACGTTCGCGGTGCCGAACC 1358

Qy        773 TCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCG 832
              || |  || || ||| |    |||||      |  |||||||||| || |  || ||||||
Db       1359 ACTACGACCCCGGCGCGTTGACCGTCCCGGTGATCGCGCTGCTGCTCGCCTCGGTGGCCG 1418

Qy        833 ACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCC 892
              ||  || |||| ||||   ||| |  |||| |  |   |||| |||||| ||||||
Db       1419 TCGGTGATGTGCGCGGCGTCCGCTCCTGGCTGGGGACCAGGACGATGGTGCTGCTGGGGG 1478

Qy        893 AGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGA 952
              |     |||||||||||||||| || |||| ||||||||||  ||||| |||||  |
Db       1479 AACTCACCTTCGCCTTCTACCTCGTGCACTACCTGATCATCCAGTACGGGCACCGCTTCG 1538

Qy        953 TGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCC 1012
              |||||  | | ||||||  | |||||||||  ||| ||||||||  ||| |||| |||
Db       1539 CCGGCGGGAAGCAGGGCTATTACCGGCAGTGGGACACACCGGCCGCCGTCGGGCTGACCC 1598

Qy       1013 TGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGC 1072
              ||      |  | ||||| |  |      |  | ||||||  | |||| ||  |||
Db       1599 TGCTCGCCTTCACGCTGGCGCTGGGGCTGTCGGCGTTCCTGCACTTCTTCGTGGAGAAGC 1658

Qy       1073 CCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG 1132
              |  ||||||      | ||  | | |  | | | | || | |    || || | ||   |
Db       1659 CGGTCATGCGAACCCTGGGACGGC-CGCGGCGGTCCCCGGACGCCGGCTCGACACCCAGG 1717

Qy       1133 CGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTAC 1192
                 ||||   |   |       |||  || | ||  |  |  | | | | ||||  |
Db       1718 TCCGAACCCGCCCCGTCCGGCACTCCGTAGCCGACGCGGGACGACCGGTGCGCGGCGCGC 1777

Qy       1193 CCGGGCCCGCGCACACCCCCGCAGCG 1218
              ||  |  ||||| |  | |||  | |
Db       1778 CCTCGGGCGCGCCCCGCACCGGTGTG 1803
```

RESULT 9
ADM45913
ID ADM45913 standard; DNA; 84428 BP.
XX
AC ADM45913;
XX
DT 03-JUN-2004 (first entry)
XX
DE Streptomyces mycarofaciens midecamycin polyketide synthetase DNA.
XX
KW macrolide; midecamycin biosynthesis; polyketide synthetase; PKS;
KW antimicrobial; ds; gene.
XX
OS Streptomyces mycarofaciens.
XX
PN JP2004049100-A.
XX
PD 19-FEB-2004.
XX
PF 19-JUL-2002; 2002JP-00210516.
XX
PR 19-JUL-2002; 2002JP-00210516.
XX
PA (MEIJ) MEIJI SEIKA KAISHA LTD.
XX
DR WPI; 2004-174115/17.
DR P-PSDB; ADM45914, ADM45915, ADM45916, ADM45917, ADM45918, ADM45919,
DR ADM45920, ADM45921, ADM45922, ADM45923, ADM45924, ADM45925, ADM45926,
DR ADM45927, ADM45928, ADM45929, ADM45930, ADM45931, ADM45932, ADM45933,
DR ADM45934, ADM45935, ADM45936, ADM45937, ADM45938, ADM45939, ADM45940,
DR ADM45941, ADM45942, ADM45943, ADM45944, ADM45945, ADM45946, ADM45947,
DR ADM45948, ADM45949, ADM45950, ADM45951, ADM45952, ADM45953, ADM45954,
DR ADM45955.
XX
PT New polynucleotides encoding proteins important in the biosynthesis of
PT midecamycin.

XX
PS Claim 2; SEQ ID NO 1; 264pp; Japanese.
XX
CC The invention relates to a novel polynucleotide, specifically a
CC biosynthesis gene, comprising a nucleotide sequence which encodes a
CC protein, where the gene is concerned with the biosynthesis of
CC midecamycin. The polynucleotide of the invention demonstrates
CC antimicrobial activity and may be useful for the biosynthesis of
CC midecamycin, as well as for manufacturing macrolide compounds other than
CC midecamycin. The current sequence is that of the Streptomyces
CC mycarofaciens midecamycin polyketide synthetase (PKS) DNA of the
CC invention.
XX
SQ Sequence 84428 BP; 11456 A; 29098 C; 30275 G; 13599 T; 0 U; 0 Other;

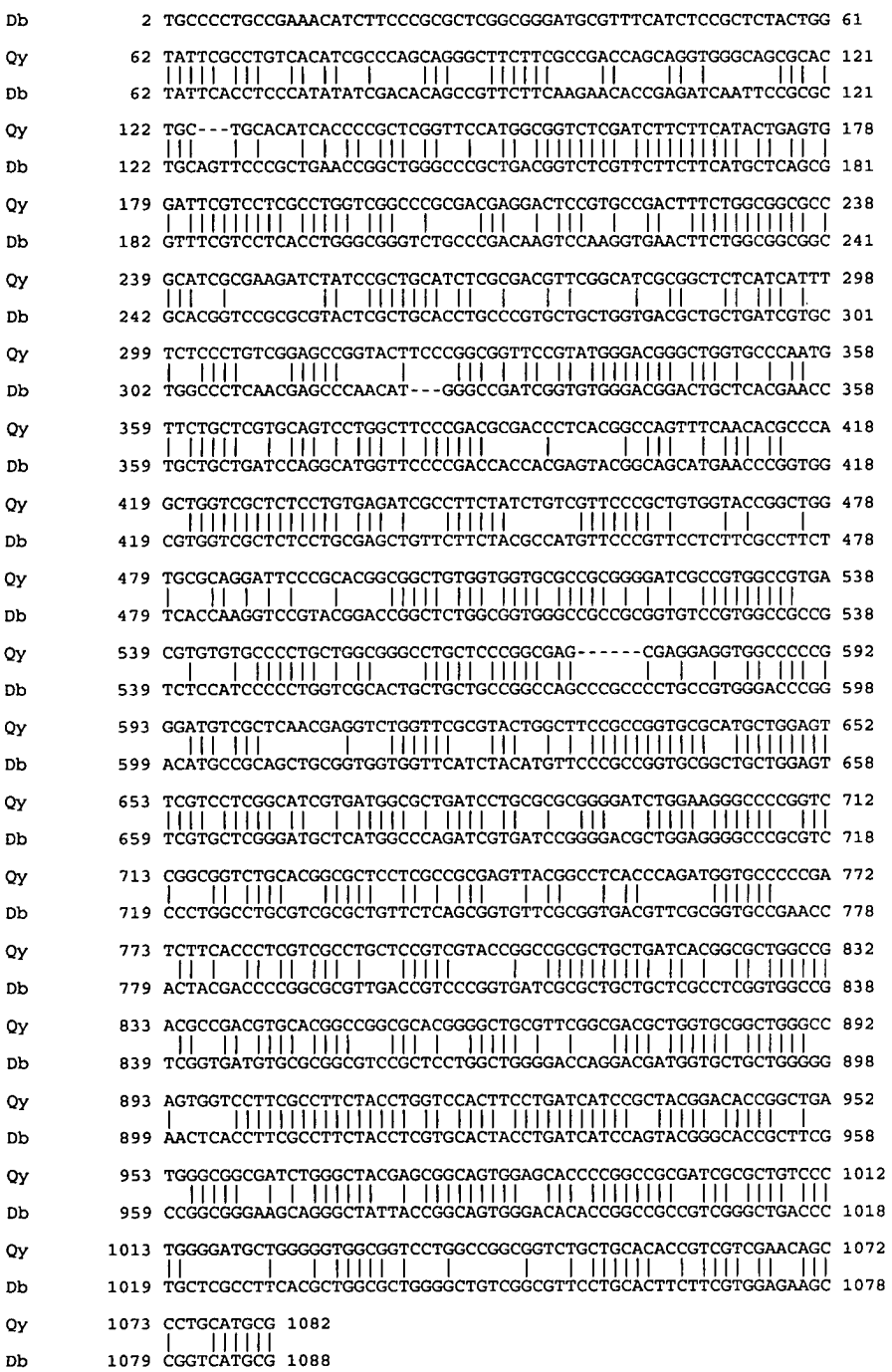
Query Match 28.6%; Score 356.4; DB 12; Length 84428;
Best Local Similarity 59.3%; Pred. No. 4.8e-48;
Matches 666; Conservative 0; Mismatches 446; Indels 12; Gaps 3;

```
Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61
              ||| |  ||       | ||| |||| |||  |||| | ||        || || |||  ||
Db      19064 TGCCCTTGCCGAAACACCTGCCGTCGCTCGGCGGCATGCGGGCCATCGCCGCACTGGTGG 19123

Qy         62 TATTCGCCTGTCACATCG---CCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCG 118
              | |||  || ||| ||||    |||||| |   |||  |  || | ||         || ||
Db      19124 TGTTCTGCTCTCATATCGCTTCCCAGCCGTTTTTCCGCAACGCCAAGATAAACTCCACCG 19183

Qy        119 CACTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTG 178
              ||| | | |   |   |   || ||  |  |||||||||| |||||||||| || || |
Db      19184 CACAGGTCCCGCTGGACGTCCTGGGGCCGCTGGCGGTCTCGTTCTTCTTCATGCTCAGCG 19243

Qy        179 GATTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCC 238
              |||||||||| ||||| ||| |     |||    |||  |||  | ||||||| ||||
Db      19244 GATTCGTCCTCACCTGGGCGGGCATGCCCGACCCGTCCAAGCCTGCCTTCTGGCGCCGCC 19303

Qy        239 GCATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTT 298
              |    |    | ||||  ||||||| ||  ||     |      || ||       ||| |
Db      19304 GTTGGGTTCGGGTCTACTCGCTGCACCTGCCGGTCCTGCTGCTCACGCTGGCGATCGTGC 19363

Qy        299 TCTCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATG 358
              | |  |||  ||| ||      |  ||||| || || ||||||||  |  |   |||
Db      19364 TGTGGCTGAAGGAACC---CAATATGGGCGGGTCGGTGTGGGACGGCTTCCTCAGCAACC 19420

Qy        359 TTCTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCA 418
              | |||||||| ||||| |||   ||||||      ||     | |||  | ||| ||
Db      19421 TGCTGCTCGTCCAGTCGTGGTGCCCCGACTACCACCAGTACGGCAGCATGAACCCGGTGG 19480

Qy        419 GCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGG 478
              |||||  |||||||| |||||     |||||      | ||||||  |   || |    |
Db      19481 CGTGGTCCCTCTCCTGCGAGATGCTGTTCTACGCCGCCTTCCCGTTCCTGTTCGCCTTCT 19540

Qy        479 TGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGA 538
              |    || ||| |  ||   ||||||||||  || | |  ||| ||| |||| | ||
Db      19541 TCTCCAAGATGCGTGCCGAGCGGCTGTGGTCCTGGGTCCTGGGCATCTCCGTCGTCGCCG 19600

Qy        539 CGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGC------CCCCG 592
              ||   ||||||    || ||     |||||||||   ||  |      | || |       |||
Db      19601 CGGCCGTGCCCGCCCTCGCCCTGCTGCTCCCCTCGGCCCCCACGCTGCCCTGGGACCCGA 19660

Qy        593 GGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGT 652
              |||  ||      | |   |||||||   |||  ||||||||||||||  ||||||| |
Db      19661 ACATGCCGGAGCTCCAATACTGGTTCATCTACATGCTTCCGCCGGTGCGGCTGCTGGAAT 19720

Qy        653 TCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712
              |||  |||||| || |||||||| |||| |   |    ||   |||||  |||||    |
Db      19721 TCGCGCTCGGCGTCCTGATGGCGCAGATCGTCAGGCGCGGCCGCTGGATCGGCCCGACCC 19780

Qy        713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772
              ||| ||| ||| ||  |||  |||||||     | ||  ||  ||       ||||  |
Db      19781 CGGGGGTGTGCGCGCTGCTGTTCGCCGGCGCGTTCGCGCTGTCCTTCGCCCTGCCGTCCT 19840

Qy        773 TCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCG 832
              |  | | ||| ||     | |||   | |  |||||||||| ||    | |||||  |
Db      19841 ATCTGGCTCGCGTAGCGCCGACGGTCCCGCTGATCGCGCTGCTGCTCGGCTCCCTGGCAG 19900

Qy        833 ACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCC 892
              | |||| | | |||    ||  || |||||  |  |   |||| ||||||  ||||||
Db      19901 CTGGCGACATACGCGGTACCCGGTCGTGGCTGGGCACCCGGACGATGGTGCTGCTGGGTG 19960

Qy        893 AGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGA 952
              |     |||||||||||||| |  |||||| |||  ||||||  |||||  |||||  |
Db      19961 AACTCACCTTCGCCTTCTACGTCATCCACTACCTCGTCATCCAGTACGGGCACCGCTTCC 20020

Qy        953 TGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCC 1012
              | ||||| || ||| ||||| | || ||||||  ||||||||||||||||  ||  ||
Db      20021 TCGGCGGTGAGCTGAGCTACTACCGACAGTGGGACACCCCGGCCGCGATCGGCCTCACCG 20080

Qy       1013 TGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGC 1072
              |       || ||| |    ||       || ||  || || |||  | |||| ||  |||
Db      20081 TTCTCGCCCTCGGGCTCAGCGTGGGCCTCGCCGCGCTCCTCCACTTCTTCGTGGAGAAGC 20140

Qy       1073 CCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCG 1116
              |   | | ||     ||||| ||  | |    || |||||  ||
Db      20141 CGGTCGTCCGGGCCCTCGGCCGCTCCGGCAAGGCGTCCCGCGCG 20184
```

RESULT 10
AAN92630
ID AAN92630 standard; DNA; 1167 BP.
XX
AC AAN92630;
XX
DT 25-MAR-2003 (revised)
DT 15-MAY-1990 (first entry)
XX
DE DNA fragment containing acyB gene.
XX
KW Macrolide; antibiotics; Streptomyces; acylation; ds.
XX
OS Streptomyces sp.
XX
PN EP345546-A.
XX
PD 13-DEC-1989.
XX
PF 24-MAY-1989; 89EP-00109425.
XX
PR 24-MAY-1988; 88JP-00125091.
PR 03-MAR-1989; 89JP-00050120.
XX
PA (SAOC) SANRAKU INC.
PA (SAOC) MERCIAN CORP.
XX
PI Arisawa A, Kawamura N, Kojima I, Okumura Y, Okamura K, Tone H;
PI Okamoto R;
XX
DR WPI; 1989-365353/50.
DR P-PSDB; AAP93671.
XX
PT DNA coding for enzyme acylating 4"-position of macrolide antibiotics -
PT contg. acyB gene and derived from Streptomyces microorganism.
XX
PS Claim 6; Page 28-33; 54pp; English.
XX
CC DNA fragment of acyB gene reading upstream to give acyB1. May be used to
CC produce a 4" acylated macrolide antibiotic without preparing and then
CC acylating it. (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 1167 BP; 146 A; 417 C; 365 G; 239 T; 0 U; 0 Other;

Query Match 28.2%; Score 351.2; DB 1; Length 1167;
Best Local Similarity 59.6%; Pred. No. 5.1e-47;
Matches 650; Conservative 0; Mismatches 428; Indels 12; Gaps 3;

```
Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61
              ||| || ||       ||||  |  | |||  |||  | || |      | || ||   ||
```



RESULT 11
AAZ35999
ID AAZ35999 standard; DNA; 1646 BP.
XX
AC AAZ35999;
XX
DT 14-FEB-2000 (first entry)
XX
DE S. kitasatoensis macrolide antibiotic acylation enzyme gene.
XX
KW Streptomyces kitasatoensis; macrolide antibiotic acylation enzyme;
KW microbe; acylating macrolide antibiotic; ds.
XX
OS Streptomyces kitasatoensis.
XX
FH Key Location/Qualifiers
FT CDS 9. .1178
FT /*tag= a
FT /product= "macrolide antibiotic acylation enzyme"
XX
PN JP11285384-A.
XX
PD 19-OCT-1999.
XX
PF 02-APR-1998; 98JP-00089651.
XX
PR 02-APR-1998; 98JP-00089651.
XX
PA (ASAH) ASAHI KASEI KOGYO KK.
XX
DR WPI; 2000-026813/03.
DR P-PSDB; AAY56305.
XX
PT A DNA containing a macrolide antibiotic acylation enzyme gene - used in
PT the expression of the protein.
XX
PS Claim 1; Page 7-9; 10pp; Japanese.
XX
CC The present sequence encodes a Streptomyces kitasatoensis macrolide
CC antibiotic acylation enzyme. Introducing the macrolide antibiotic
CC acylating enzyme gene into a microbe strain can improve its acylating
CC macrolide antibiotic-producing titre. A method from the present invention
CC can be used for expressing the encoded protein
XX
SQ Sequence 1646 BP; 200 A; 604 C; 551 G; 291 T; 0 U; 0 Other;

Query Match 27.7%; Score 345; DB 3; Length 1646;
Best Local Similarity 58.4%; Pred. No. 4.7e-46;
Matches 682; Conservative 0; Mismatches 470; Indels 15; Gaps 4;

```
Db        310 TGTGGCTCCAGGAACCATCCATG---GGCCGGTCCGTGTGGGACGGGTTCCTGACGAATC 366

Qy        359 TTCTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCA 418

Db        367 TCTTCCTGGTCCAGGCGTGGTTCCCGGACTTCCACGAATACGCCAGCATGAATCCGGTGG 426

Qy        419 GCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGG 478

Db        427 CGTGGTCGCTCTCCGCCGAGATGCTCTTCTACGCGTCGTTCCCGTTCCTGTTCGCCGCAC 486

Qy        479 TGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGA 538

Db        487 TGTCCAAGGTGCGCGTCGAGCGGCTGTGGAAGTGGGCCGCCGGCCTGGCCGTCGCCGTCG 546

Qy        539 CGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGC------CCCCG 592

Db        547 TGGCGGTGCCGTCCGTCGCCCTGCTGATGCCGTCCGACCCGGCGCTGCCGTGGGACGCCG 606

Qy        593 GGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGT 652

Db        607 GGATGCCGGAGATGGAGTACTGGTTCATCTATGTGTTCCCGCCGATCCGGCTCATGGAGT 666

Qy        653 TCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712

Db        667 TCACCCTCGGCATCGTCATGGCGCTGATCGTGAAGAACGGCCGCTGGGCGGGCCCCAAGC 726

Qy        713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772

Db        727 CCCTCGTCTGCGCCCTGGCGTTCGGCGCGCTGTACGGGATCTC---GTTCGCGGTGCCGG 783

Qy        773 TCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCG 832

Db        784 ACTGGCTGGGCCGCAGCCTGGCGGTCCCGGCGGTCGCCCTGCTGCTGGGCTCGCTCGCCG 843

Qy        833 ACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCC 892

Db        844 CGGGCGACGTGCGCGGCGCCCGCTCGCCGCTGGCCTCGAAGACGATGGTCCTGCTCGGCG 903

Qy        893 AGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGA 952

Db        904 AACTCACCTTCGCCTTCTACCTCGTCCACTACCTGGTCATCCAGTACGGGCACCGTTTCC 963

Qy        953 TGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCC 1012

Db        964 TCGGCGGTGAGCTGAGCTACTACCGCCAGTGGGACACCCCGGCCGCGGCGGGCCTCGTGG 1023

Qy       1013 TGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGC 1072

Db       1024 TGCTCGCCTTCGCGCTGAGCCTCGCCCTCGCCGCCTTCCTCCACTTCTTCGTGGAGAAGC 1083

Qy       1073 CCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG 1132

Db       1084 CCGTCATGCGGACCATCGGCCGGCCGCGCCGCCCGGCCCCGCCCGCCCCGGACCCGGCCC 1143

Qy       1133 CGGCTCCCCGGAACTCACCCGCGGCCG 1159

Db       1144 CGCGGTCCGGCCCGCCGGCCGCCGCCG 1170


RESULT 12
AAN92629/c
ID   AAN92629 standard; DNA; 2729 BP.
XX
AC   AAN92629;
XX
DT   25-MAR-2003  (revised)
DT   15-MAY-1990  (first entry)
XX
DE   DNA fragment containing acyB gene.
XX
KW   Macrolide; antibiotics; Streptomyces; acylation; ds.
XX

OS   Streptomyces sp.
XX
PN   EP345546-A.
XX
PD   13-DEC-1989.
XX
PF   24-MAY-1989;   89EP-00109425.
XX
PR   24-MAY-1988;   88JP-00125091.
PR   03-MAR-1989;   89JP-00050120.
XX
PA   (SAOC ) SANRAKU INC.
PA   (SAOC ) MERCIAN CORP.
XX
PI   Arisawa A,  Kawamura N,  Kojima I,  Okumura Y,  Okamura K,  Tone H;
PI   Okamoto R;
XX
DR   WPI; 1989-365353/50.
XX
PT   DNA coding for enzyme acylating 4"-position of macrolide antibiotics -
PT   contg. acyB gene and derived from Streptomyces microorganism.
XX
PS   Claim 7; Page 28-33; 54pp; English.
XX
CC   DNA fragment contains two open reading frames, one upstream, one down-
CC   stream, coding for acyB1 and acyB2 genes respectively. (Updated on 25-MAR
CC   -2003 to correct PA field.)
XX
SQ   Sequence 2729 BP; 475 A; 891 C; 962 G; 401 T; 0 U; 0 Other;

  Query Match             25.8%;  Score 321.6;  DB 1;  Length 2729;
  Best Local Similarity   59.0%;  Pred. No. 2.3e-42;
  Matches  629;  Conservative    0;  Mismatches  424;  Indels   13;  Gaps    4;

Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61

Db       1224 TGCCCCTGCCGAAACATCTTCCCGCGCTCGGCGGGATGCGTTTCATCTCCGCTCTACTGG 1165

Qy         62 TATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCAC 121

Db       1164 TATTCACCTCCCATATATCGACACAGCCGTTCTTCAAGAACACCGAGATCAATTCCGCGC 1105

Qy        122 TGC---TGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTG 178

Db       1104 TGCAGTTCCCGCTGAACCGGCTGGGCCCGCTGACGGTCTCGTTCTTCTTCATGCTCAGCG 1045

Qy        179 GATTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCC 238

Db       1044 GTTTCGTCCTCACCTGGGCGGGTCTGCCCGACAAGTCCAAGGTGAACTTCTGGCGGCGGC 985

Qy        239 GCATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTT 298

Db        984 GCACGGTCCGCGCGTACTCGCTGCACCTGCCCGTGCTGCTGGTGACGCTGCTGATCGTGC 925

Qy        299 TCTCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATG 358

Db        924 TGGCCCTCAACGAGCCCAACAT---GGGCCGATCGGTGTGGGACGGACTGCTCACGAACC 868

Qy        359 TTCTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCA 418

Db        867 TGCTGCTGATCCAGGCATGGTTCCCCGACCACCACGAGTACGGCAGCATGAACCCGGTGG 808

Qy        419 GCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGG 478

Db        807 CGTGGTCGCTCTCCTGCGAGCTGTTCTTCTACGCCATGTTCCCGTTCCTCTTCGCCTTCT 748

Qy        479 TGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGA 538

Db        747 TCACCAAGGTCCGTACGGACCGGCTCTGGCGGTGGGCCGCCGCGGTGTCCGTGGCCGCCG 688

Qy        539 CGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAG------CGAGGAGGTGGCCCCCG 592

Db        687 TCTCCATCCCCCTGGTCGCACTGCTGCTGCCGGCCAGCCCGCCCCTGCCGTGGGACCCGG 628
```

```
Qy        593 GGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGT 652
              ||| |||       |   ||||||   |||  | | ||||||||||  |||||||||
Db        627 ACATGCCGCAGCTGCGGTGGTGGTTCATCTACATGTTCCCGCCGGTGCGGCTGCTGGAGT 568

Qy        653 TCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712
              |||| ||||| ||  | |||||| | |||| ||  |  |||   |||||  ||||||  |||
Db        567 TCGTGCTCGGGATGCTCATGGCCCAGATCGTGATCCGGGGACGCTGGAGGGGCCCGCGTC 508

Qy        713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772
              |   || ||||   |||||  || | |||   | ||   | ||      ||||||
Db        507 CCCTGGCCTGCGTCGCGCTGTTCTCAGCGGTGTTCGCGGTGACGTTCGCGGTGCCGAACC 448

Qy        773 TCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCG 832
              || |  || || ||| |   |||||      |  |||||||||| || |  || ||||||
Db        447 ACTACGACCCCGGCGCGTTGACCGTCCCGGTGATCGCGCTGCTGCTCGCCTCGGTGGCCG 388

Qy        833 ACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCC 892
              ||  || |||| ||||   ||| |  ||||| |  |   |||| |||||| ||||||
Db        387 TCGGTGATGTGCGCGGCGTCCGCTCCTGGCTGGGGACCAGGACGATGGTGCTGCTGGGGG 328

Qy        893 AGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGA 952
              |      |||||||||||||||| || |||| ||||||||||  |||||  |||||   |
Db        327 AACTCACCTTCGCCTTCTACCTCGTGCACTACCTGATCATCCAGTACGGGCACCGCTTCG 268

Qy        953 TGGGCGGCGATCTGGGCTACGAGCGGC-AGTGGAGCACCCCGGCCGCGATCGCGCTGTCC 1011
              |||||  | | ||||||  | ||||    ||   |||| |  |||  || ||||| |
Db        267 CCGGCGGGAAGCAGGGCTATTACCGGCGTCGGGCTGACCCTGCTCGCCTTCACGCTGGCG 208

Qy       1012 CTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACA 1057
              |||||| ||  || | |   |  | |   ||  |    || |  ||
Db        207 CTGGGGCTGTCGGCGTTCCTGCACTTCTTCGTGGAGAAGCCGGTCA 162


RESULT 13
AAD47222/c
ID   AAD47222 standard; DNA; 19016 BP.
XX
AC   AAD47222;
XX
DT   24-FEB-2003  (first entry)
XX
DE   Streptococcus sp. H021 DNA fragment for rabelomycin biosynthetic pathway.
XX
KW   Gene cluster; rabelomycin biosynthetic pathway; angucycline compound;
KW   drug screening; enzyme; gene; ds.
XX
OS   Streptococcus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(1. .561)
FT                   /*tag=  m
FT                   /product= "OrfS"
FT                   /note= "No stop codon"
FT                   /partial
FT   CDS             complement(580. .1335)
FT                   /*tag=  n
FT                   /product= "OrfR"
FT   CDS             complement(1391. .2701)
FT                   /*tag=  l
FT                   /product= "OrfQ"
FT                   /transl_except= (pos:2699. .2701, aa:Met)
FT   CDS             complement(2712. .3710)
FT                   /*tag=  k
FT                   /product= "OrfH"
FT   CDS             complement(3793. .4848)
FT                   /*tag=  s
FT                   /product= "Orf2"
FT   CDS             complement(4854. .5438)
FT                   /*tag=  j
FT                   /product= "OrfO"
FT   CDS             complement(5494. .6687)
FT                   /*tag=  o
FT                   /product= "OrfY"
FT   CDS             complement(6780. .8051)
FT                   /*tag=  q
FT                   /product= "OrfJ"
FT   CDS             complement(8178. .8939)
FT                   /*tag=  i
FT                   /product= "OrfV"
FT                   /transl_except= (pos:8937. .8939, aa:Met)
FT   CDS             complement(9014. .10555)
FT                   /*tag=  h
FT                   /product= "OrfM"
FT                   /transl_except= (pos:10553. .10555, aa:Met)
FT   CDS             complement(10557. .11504)
FT                   /*tag=  g
FT                   /product= "OrfL"
FT   CDS             complement(11570. .12352)
FT                   /*tag=  d
FT                   /product= "OrfD"
FT                   /transl_except= (pos:12350. .12352, aa:Met)
FT   CDS             complement(12421. .12684)
FT                   /*tag=  c
FT                   /product= "OrfC"
FT   CDS             complement(12681. .13910)
FT                   /*tag=  b
FT                   /product= "OrfB"
FT   CDS             complement(13907. .15142)
FT                   /*tag=  a
FT                   /product= "OrfA"
FT                   /transl_except= (pos:15140. .15142, aa:Met)
FT   CDS             complement(15208. .15537)
FT                   /*tag=  f
FT                   /product= "OrfF"
FT   CDS             complement(15586. .17055)
FT                   /*tag=  e
FT                   /product= "OrfE"
FT   CDS             17692. .18492
FT                   /*tag=  r
FT                   /product= "Orf1"
FT   CDS             complement(18603. .19016)
FT                   /*tag=  p
FT                   /product= "OrfR1"
FT                   /note= "No start codon"
FT                   /partial
XX
PN   WO200274800-A1.
XX
PD   26-SEP-2002.
XX
PF   15-MAR-2002; 2002WO-FI000214.
XX
PR   19-MAR-2001; 2001FI-00000553.
XX
PA   (GALI-) GALILAEUS OY.
XX
PI   Palmu K,  Kunnari T;
XX
DR   WPI; 2003-018791/01.
DR   P-PSDB; AAE29801, AAE29802, AAE29803, AAE29804, AAE29805, AAE29806,
DR   AAE29807, AAE29808, AAE29809, AAE29810, AAE29811, AAE29812, AAE29813,
DR   AAE29814, AAE29815, AAE29816, AAE29817, AAE29818, AAE29819.
XX
PT   Novel DNA fragment for producing hybrid compounds, is a gene cluster for
PT   rabelomycin biosynthetic pathway of Streptomyces and is included in
PT   specific PstI fragments of Streptomyces species genome.
XX
PS   Claim 2; Page 29-37; 66pp; English.
XX
CC   The invention relates to an isolated and purified DNA fragment, which is
CC   the gene cluster for rabelomycin biosynthetic pathway of Streptomyces
CC   bacteria, being included in two 9.5 kb flanked PstI fragments of
CC   Streptomyces sp. genome. The method of the invention is useful for
CC   production of hybrid compounds, by transferring the DNA fragment or at
CC   least one of the genes selected from open reading frames A,B,C,D,E,F,L,M,
```

```
CC   V,O,H,Q,R,Y,J,1 and 2 derived from the DNA fragment, into a Streptomyces
CC   host (e.g. S. lividans, S. argillaceus or S. galilaeus), cultivating the
CC   recombinant strain obtained and isolating the angucycline compounds
CC   produced. The above production method is useful for generating novel
CC   compounds for drug screening. The present sequence is Streptococcus sp.
CC   H021 DNA fragment which is involved in rabelomycin biosynthetic pathway
XX
SQ   Sequence 19016 BP; 2637 A; 6421 C; 7187 G; 2771 T; 0 U; 0 Other;

  Query Match             20.2%;  Score 251.4;  DB 8;  Length 19016;
  Best Local Similarity   53.8%;  Pred. No. 2.6e-31;
  Matches  659;  Conservative    0;  Mismatches  521;  Indels   45;  Gaps    5;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
Db       6660 CTCCCCTCGCTCACCGGGATGCGTTTTCTCGCCGCCCTGCTGGTGTTCTCCGTGCACGGG 6601

Qy         79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCAC---ATCACC 135
Db       6600 GCCGCCGCCGGCGTCTTCAAGGACCAGGGGGTCGCAGCCGACTACTACCGCTGGTTCGGC 6541

Qy        136 CCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGG 195
Db       6540 AACGCCGGCGCGGTCGGCGTCTCCTTCTTCTTCATGCTCAGCGGCTTCGTCCTCACCTGG 6481

Qy        196 TCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTAT 255
Db       6480 TCGGTGCGTCCCGCCGACACCGTCCGCGGATTCTGGCGGCGGCGCCTGCTGAAGATCTTC 6421

Qy        256 CCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCG 315
Db       6420 CCCAACCACCTGGTGACGTTCGTCGTGGC----------------CATCGTCCTGCTGACG 6376

Qy        316 GTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCC 375
Db       6375 GTGACCAGTACCGCGGTCGCCTTCCCCGAGACGCTCGCCAACCTGTTCCTGGTGCACGCC 6316

Qy        376 TGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGT 435
Db       6315 TGGGTGCCCGACAGCGGCTACGTCGAAACGGCCAACACCGTCAGCTGGTCGCTCTCCGTG 6256

Qy        436 GAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCA 495
Db       6255 GAGCTCCTCTTCTACCTGTCCTTTCCGCTGCTCATCAAGGGCGTGAGCCGGATCGCGCCG 6196

Qy        496 CGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTG 555
Db       6195 GCCGTGCTCTGGTACGCGGCGGGTGGCGTCGTGCTGGTCATCATGCTGATGCCGCTGATC 6136

Qy        556 GCGGGCCTGCTCCCGGCGAGCGAGGAGG------TGGCCCCCGGGATGTCGCTCAACGAG 609
Db       6135 GCGCAGGTGCTGCTGCCGGACACCCCGGGCTTCATGTTCATGAAGATCTCCTGGACCCAG 6076

Qy        610 GTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTG 669
Db       6075 ATCTGGTTCGTGTACGTCTTCCCCGTCGCCCGGCTCCTCGAATTCGTGCTCGGCATGCTG 6016

Qy        670 ATGGCGCTGATCCTGC------GCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGC 723
Db       6015 CTGGCGCGCATCGTGCTGTCCGGCCGGTGGATCGGGCTCGGCGTGATACCGCCGGCCCTC 5956

Qy        724 ACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTC 783
Db       5955 CTGACGATCGCCGCCTACGTGGGGGCCGTACTGATCGACCACAACCCGCTCTACAACTAC 5896

Qy        784 GTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTG 843
Db       5895 GTGGCGATCACCGTCGTCCCGCTGGCCCTGCTGATCCCGGCCGCCGCCGCCTCCGACGCC 5836

Qy        844 CACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTC 903
Db       5835 AAGGGCCGCGAGAGCTTGCTGAGCAAGCGCCCCATGGTCTGGCTCGGGGAGATCTCGTTC 5776

Qy        904 GCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGAT 963
Db       5775 GCCTTCTACTGCGTGCACTATCTGATCCTGATCTACGGTCACCGGCTGTTCGGCTCGGAC 5716

Qy        964 C---------------TGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTG 1008
Db       5715 CCCAACATCTTCGGCAAGCCCTCCGGACCCGCCTGGAGCACACCCGGCGGCCTGCTCTTC 5656

Qy       1009 TCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAA 1068
Db       5655 CTGGCGGCGGCCCTCGTCGTGTCCGTGCTCGCCGCGTGGGCCCTGTACGCGATCGTCGAA 5596

Qy       1069 CAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCC 1128
Db       5595 CGCCCCGTCATGCGCCGCTGGGGCCGCCCGCCGGGACCCCGGCGCACCGAGCAGACGGCC 5536

Qy       1129 ACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGC 1188
Db       5535 GTCCGGCCGGAACGGGCCGCCGAAGCCCCCGCGGCCCACTGAGGACGCGGCCCCCGACCA 5476

Qy       1189 GTACCCGGGCCCGCGCACACCCCCG 1213
Db       5475 CCGACCGCACCGCCCCACACACCCG 5451


RESULT 14
ADE86070
ID   ADE86070 standard; DNA; 47988 BP.
XX
AC   ADE86070;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Streptomyces hygroscopicus non-ribosomal peptide synthetase complex DNA.
XX
KW   Peptide synthetase; enzyme; antibiotic; AC98; gene; ss.
XX
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             77. .1048
FT                   /*tag=  a
FT                   /product= "Acetyltransferase"
FT                   /note= "ORF1, specifically claimed in Claim 30, encodes
FT                   SEQ ID NO:21"
FT   CDS             1045. .2460
FT                   /*tag=  b
FT                   /product= "ABC transporter"
FT                   /note= "ORF2, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:22"
FT   CDS             2495. .3406
FT                   /*tag=  c
FT                   /product= "ABC transporter"
FT                   /note= "ORF3, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:23"
FT   CDS             3403. .4293
FT                   /*tag=  d
FT                   /product= "ABC transporter"
FT                   /note= "ORF4, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:24"
FT   CDS             4359. .5636
FT                   /*tag=  e
FT                   /product= "Mannosyltransferase"
FT                   /note= "ORF5, specifically claimed in Claim 31, encodes
FT                   SEQ ID NO:25"
FT   CDS             5822. .7234
FT                   /*tag=  f
FT                   /product= "Unknown"
FT                   /note= "ORF6, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:26"
FT   CDS             7293. .8822
FT                   /*tag=  g
FT                   /product= "Unknown"
```

```
FT                   /note= "ORF7, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:27"
FT   CDS             9012. .10025
FT                   /*tag=  h
FT                   /product= "Methyltransferase"
FT                   /note= "ORF8, specifically claimed in Claim 32, encodes
FT                   SEQ ID NO:28"
FT   CDS             10069. .18309
FT                   /*tag=  i
FT                   /product= "Non-ribosomal peptide synthetase"
FT                   /gene= "mppA"
FT                   /note= "The CDS is specifically claimed in Claim 2,
FT                   encodes SEQ ID NO:2"
FT   CDS             18309. .29312
FT                   /*tag=  j
FT                   /product= "Non-ribosomal peptide synthetase"
FT                   /gene= "mppB"
FT                   /note= "The CDS is specifically claimed in Claim 3,
FT                   encodes SEQ ID NO:4"
FT   CDS             29319. .30628
FT                   /*tag=  k
FT                   /product= "D-aminoacyl hydrolase superfamily"
FT                   /note= "ORF9, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:29"
FT   CDS             30658. .32010
FT                   /*tag=  l
FT                   /product= "Efflux protein"
FT                   /note= "ORF10, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:30"
FT   CDS             32181. .33407
FT                   /*tag=  m
FT                   /product= "Isovaleryl transferase"
FT                   /note= "ORF11, specifically claimed in Claim 33, encodes
FT                   SEQ ID NO:31"
FT   CDS             33422. .34792
FT                   /*tag=  n
FT                   /product= "Isovaleryl transferase"
FT                   /note= "ORF12, specifically claimed in Claim 34, encodes
FT                   SEQ ID NO:32"
FT   CDS             34905. .35930
FT                   /*tag=  o
FT                   /product= "Arginine cyclase"
FT                   /note= "ORF13, specifically claimed in Claim 35, encodes
FT                   SEQ ID NO:33"
FT   CDS             36383. .37267
FT                   /*tag=  p
FT                   /product= "Transaminase"
FT                   /note= "ORF14, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:35"
FT   CDS             37264. .38511
FT                   /*tag=  q
FT                   /product= "Transaminase"
FT                   /note= "ORF15, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:37"
FT   CDS             38544. .40378
FT                   /*tag=  r
FT                   /product= "Transaminase"
FT                   /transl_except= (pos:39291. .39292,aa:Arg)
FT                   /note= "ORF16, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:39; this codon has an apparent 1 nucleotide
FT                   deletion which alteres the reading frame"
FT   CDS             40444. .40659
FT                   /*tag=  s
FT                   /product= "Cda-ORFX homologue"
FT                   /note= "ORF17, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:41"
FT   CDS             complement(41100. .42554)
FT                   /*tag=  t
FT                   /product= "Putative secreted protein"
FT                   /note= "ORF18, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:43"
FT   CDS             complement(42936. .44648)
FT                   /*tag=  u
FT                   /product= "Hypothetical protein"
FT                   /note= "ORF19, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:45"
FT   CDS             complement(44581. .45435)
FT                   /*tag=  v
FT                   /product= "ABC transporter"
FT                   /note= "ORF20, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:47"
FT   CDS             complement(45504. .45920)
FT                   /*tag=  w
FT                   /product= "Putative lipoprotein"
FT                   /note= "ORF21, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:49"
FT   CDS             complement(45991. .47181)
FT                   /*tag=  x
FT                   /product= "Two component sensor kinase"
FT                   /note= "ORF22, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:51"
FT   CDS             complement(47308. .47988)
FT                   /*tag=  y
FT                   /product= "Two component response regulator"
FT                   /note= "ORF23, specifically claimed in Claim 6, encodes
FT                   SEQ ID NO:53"
XX
PN   WO2003082909-A1.
XX
PD   09-OCT-2003.
XX
PF   28-MAR-2003; 2003WO-US009746.
XX
PR   29-MAR-2002; 2002US-0368713P.
XX
PA   (AMHP ) WYETH.
XX
PI   Hucul JA,  Macgarvey N,  Greenstein M;
XX
DR   WPI; 2003-812536/76.
DR   P-PSDB; ADE86073, ADE86074, ADE86075, ADE86076, ADE86077, ADE86078,
DR   ADE86079, ADE86080, ADE86081, ADE86082, ADE86083, ADE86084, ADE86085,
DR   ADE86085, ADE86086, ADE86087, ADE86088, ADE86089, ADE86090, ADE86091,
DR   ADE86092, ADE86093, ADE86094, ADE86095.
XX
PT   New non-ribosomal peptide synthetase, useful for preparing antibacterial
PT   peptides, derived from Streptomyces, also related nucleic acid and
PT   modified antibiotics.
XX
PS   Claim 1; SEQ ID NO 1; 163pp; English.
XX
CC   The present sequence, from Streptomyces hygroscopicus NS17 (NRRL 30439)
CC   comprises open reading frames (ORFs) mppA and mppB encoding the non-
CC   ribosomal peptide synthetase (NRPS) responsible for the production of
CC   cyclic peptide antibiotic AC98, and 23 associated ORFs predicted to play
CC   a role in the biosynthesis of AC98. Each ORF is specifically claimed. The
CC   sequence is contained within cosmid pNWA117, which was isolated from a
CC   cosmid library of NS17 by screening with a 3 kb peptide synthetase
CC   fragment. The NRPS complex exists as 2 separate components, mppA and
CC   mppB, both involved in the synthesis of the core AC98. MppA ADE86071 is
CC   composed of 3 minimal modules, where each module is comprised of an
CC   adenylation, thiolation and condensation domain. MppB ADE86072 is
CC   comprised of 2 complete minimal modules and an additional condensation
CC   and thiolation domain. Sequence comparisons of ORF1-23 indicate that the
CC   encoded proteins ADE86073-ADE86085, ADE86086, ADE86087, ADE86088,
CC   ADE86089, ADE86090, ADE86091, ADE86092, ADE86093, ADE86094 and ADE86095
CC   are tailoring enzymes involved in modification of the peptide core, such
CC   as glycosylation, methylation and acylation, or are enzymes involved in
CC   resistance. The invention provides a method of producing NRPS using
CC   transformed host cells. It also provides a method of producing a cyclic
CC   peptide antibiotic, such as AC98, using the NRPS comprised of mppA and
CC   mppB. The antibiotic is preferably a lipoglycopeptide antibiotic with
CC   activity against Gram-positive pathogens. Also provided are methods of
CC   modifying NRPS, e.g. by replacement of an adenylation domain, to produce
CC   an antibiotic having a modified peptide core, and a method for evaluating
CC   the structural regions of the modified peptide.
```

```
XX
SQ   Sequence 47988 BP; 6450 A; 17292 C; 17125 G; 7121 T; 0 U; 0 Other;

  Query Match             19.8%;  Score 246.6;  DB 10;  Length 47988;
  Best Local Similarity   55.0%;  Pred. No. 1.4e-30;
  Matches  611;  Conservative    0;  Mismatches  464;  Indels   36;  Gaps    5;

Qy         17 ATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACA 76
              | |||    || || |||||  | || |      |||  |||| |||| |||  |    ||
Db      33645 ACCTGAGATCCCTGACCGGTTTGCGGTTCGTGGCGATGCTGCCGGTCTTCCTCACCCATG 33704

Qy         77 TCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCC 136
                ||      ||||| |||||  ||||  | |||||| || | |    | ||  ||    |
Db      33705 CGGCGTTCGAGGGCGTCTTCAGCGACGCGAAGGTGAGCTGGGGCTTCCTCGACGCGATGG 33764

Qy        137 CGCT---CGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCT 193
              |      ||| |  ||||| |||||| |||||||| | || || || || ||  || | |
Db      33765 GGAGCACCGGCTATATGGCCGTCTCGTTCTTCTTCGTGCTCAGCGGCTTTGTGATCACGT 33824

Qy        194 GGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCT 253
              |||||  |||    |   ||| ||| ||  |  |||||||| || ||| ||     | |||
Db      33825 GGTCGTACCGGCCCACCGACACCGCGCGCAAGTTCTGGCGCCGGCGCTTCTTCCGGGTCT 33884

Qy        254 ATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGC 313
                ||    ||  ||| ||| |  |   ||||                ||   |||   |
Db      33885 TCCCCAACCACGTCGTGACCTATGCGCTCGC-----------CCTCGGGCTGATCGCTG 33932

Qy        314 CGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGT 373
              ||||    |   | |   |||| | |     ||| || |  | | ||||| ||||
Db      33933 CGGTGGGGCTGAGTGTCGGCGTACTGCCC---TCGGTCACCCAGCTCTTCCTCGTCCAGT 33989

Qy        374 CCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCT 433
              ||||| | |||||| || |  |||| | ||     |||||          |||||||||| |
Db      33990 CCTGGGTGCCCGACCCGGCGTTCACCGACACCGGCAACAGCGTGAGCTGGTCGCTCGCGG 34049

Qy        434 GTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCG 493
                ||  | |  |||||    |   |||||| ||  |  |  |||||||  || |||
Db      34050 TGGATGTGGTGTTCTACGCGCTCTTCCCGGTGCTGCTCACGCTGGTGAACAAGATCAAGC 34109

Qy        494 CACGGCGGCTGTGGTGGTGCGCCG------CGGGGATCGCCGTGGCCGTGACGTGTGTGC 547
              |    ||||| ||||  || | ||       | |  ||||  |||||||||       |
Db      34110 CGAATCGGCTCTGGTACTGGGTCGGTGGCTCCGTCATCGGTGTGGCCGTGGTACCGGCCA 34169

Qy        548 CCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACG 607
              |  ||| || |  || |       |   |||| |  |  |  |||| ||||  ||| |
Db      34170 TCGCGCTCGCCGCGCTCCCGAGCACCCCCGAGATGCCGCTCGGCGGGGTGTCCGTCAGCC 34229

Qy        608 AGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCG 667
              ||  |||||||  | ||||    | ||||    | ||  |||||||||   || |||||||
Db      34230 AGTACTGGTTCACCTACTTCTTCCCGCTCTTCCGGCTGCTGGAGTGTGTGCTCGGCATGC 34289

Qy        668 TGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGG 727
              ||||||||| |||| |||   | || |  ||||   |||   |   |  |  |  | | |
Db      34290 TGATGGCGCGGATCGTGCTGTCCGGCAAGTGGATACGCCTGCGGGTGCTGCCCGCCGCCG 34349

Qy        728 CGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCG 787
                ||| | |         |||  | || |||||  ||| ||      ||| |   ||    |
Db      34350 TCCTCGTGGTGATCGCGTACTACTTCGCCCAGCAGGTCCCGTACCTCTACCGGCTGAGTG 34409

Qy        788 CCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACG 847
              |      | ||  | |||| ||||||||||    |||||   || |  || |||    | |
Db      34410 CGGTGACGGTGCTGCCGGTCGCGCTGCTGACGGCGGCGGCCGCGGTGGCGGACTCCGAGG 34469

Qy        848 GCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCT 907
              ||||| ||||   | | |    |   ||||  || | ||| |    |||||||||
Db      34470 GCCGGGGCACCCTGTTCGGCAGCAAGGTCATGGTCTGGTTCGGCGAACTCTCCTTCGCCT 34529

Qy        908 TCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGG 967
              |||||||| | |||  |||  || |    ||||| || | |||| | |||  |     ||
Db      34530 TCTACCTGCTGCACAACCTCGTCCTGAAGTACGGCCATCTGCTGCTCGGCCACACCGAGG 34589

Qy        968 GCTACGAGCGGCAG------------TGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGG 1015
                | |    || |            ||| ||   || |  |  ||||| |||  |   |
Db      34590 AGGAGGGCGAGCTGGTGGGCCACACCTGGGGCGTGCCCGAGGGAATCGCCCTGATCGCCG 34649

Qy       1016 GGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCT 1075
                | | |||| |  | |||   | | | ||||||||||| ||  || ||  |||
Db      34650 CCGCCTTCGCGGTGTCCCTGCTGCTGGCCTGGCTGCTGCACAACGGAGTGGAGAAGCAGG 34709

Qy       1076 GCATGCGCCTGTTCGGCAGCCGCAGGTCCGC 1106
                |||||||  |    | | |||| | | ||
Db      34710 CGATGCGCCGCTGGTCCCGACGCAAGCCGGC 34740


RESULT 15
AAF30757
ID   AAF30757 standard; DNA; 47981 BP.
XX
AC   AAF30757;
XX
DT   21-JUN-2001  (first entry)
XX
DE   Micromonospora megalomicea megalomicin biosynthetic gene cluster.
XX
KW   Megalomicin; meg gene; polyketide synthase; antibiotic; motilide;
KW   antiparasitic; ds.
XX
OS   Micromonospora megalomicea subsp. nigra.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement(1. .144)
FT                   /*tag=  a
FT                   /partial
FT                   /gene= ""megBVI(megT)"T" /product="TDP-4-keto-6-
FT                   deoxyglucose-"
FT                   /note= "encodes AAB82201"
FT   CDS             928. .2061
FT                   /*tag=  b
FT                   /gene= "megDVI"
FT                   /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
FT                   /note= "encodes AAB82202"
FT   CDS             2072. .3382
FT                   /*tag=  c
FT                   /gene= "megDI"
FT                   /product= "TDP-megosamine glycosyltransferase"
FT                   /note= "eryCIII homologue; encodes AAB802203"
FT   CDS             3462. .4634
FT                   /*tag=  d
FT                   /gene= "megY"
FT                   /product= "mycarose O-acyltransferase"
FT                   /note= "encodes AAB82204"
FT   CDS             4651. .5775
FT                   /*tag=  e
FT                   /gene= "megDII"
FT                   /product= "TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
FT                   /note= "eryCI, DnrJ homologue, encodes AAB82205"
FT   CDS             5822. .6595
FT                   /*tag=  f
FT                   /gene= "megDIII"
FT                   /product= "daunosaminyl-N,N-dimethyltransferase"
FT                   /note= "eryCVI homologue; encodes AAB82206"
FT   CDS             6592. .7197
FT                   /*tag=  g
FT                   /gene= "megDIV"
FT                   /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
FT                   /note= "eryBVIII, dnmU homologue, encodes AAB82207"
FT   CDS             7220. .8206
FT                   /*tag=  h
FT                   /gene= "megDV"
FT                   /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT                   /note= "eryVIV, dnmV homolgoue; encodes AAB82208"
FT   CDS             complement(8228. .9220)
FT                   /*tag=  i
FT                   /gene= "megBII-1(megDVII)"
FT                   /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
FT                   /note= "encodes AAB82209"
```

```
FT   CDS             complement(9226. .10479)
FT                   /*tag= j
FT                   /gene= "megBV"
FT                   /product= "TDP-mycarose glycosyltransferase"
FT                   /note= "encodes AAB82210"
FT   CDS             complement(10483. .11424)
FT                   /*tag= k
FT                   /gene= "megBIV"
FT                   /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT                   /note= "encodes AAB82211"
FT   CDS             12181. .22821
FT                   /*tag= l
FT                   /gene= "megA"
FT                   /product= "megalomicin 6-deoxyerthronolide B synthase 1"
FT                   /note= "polyketide synthase; encodes AAB82212"
FT   misc_feature    12505. .13470
FT                   /*tag= m
FT                   /gene= "megA"
FT                   /function= "AT-L"
FT   misc_feature    13576. .13791
FT                   /*tag= n
FT                   /gene= "megA"
FT                   /function= "ACP-L"
FT   misc_feature    13849. .15126
FT                   /*tag= o
FT                   /gene= "megA"
FT                   /function= "KS1"
FT   misc_feature    15427. .16476
FT                   /*tag= p
FT                   /gene= "megA"
FT                   /function= "AT1"
FT   misc_feature    17155. .17694
FT                   /*tag= q
FT                   /gene= "megA"
FT                   /function= "KR1"
FT   misc_feature    17947. .18207
FT                   /*tag= r
FT                   /gene= "megA"
FT                   /function= "ACP1"
FT   misc_feature    18268. .19548
FT                   /*tag= s
FT                   /gene= "megA"
FT                   /function= "KS2"
FT   misc_feature    19876. .20910
FT                   /*tag= t
FT                   /gene= "megA"
FT                   /function= "AT2"
FT   misc_feature    21517. .22053
FT                   /*tag= u
FT                   /gene= "megA"
FT                   /function= "KR2"
FT   misc_feature    22318. .22575
FT                   /*tag= v
FT                   /gene= "megA"
FT                   /function= "ACP2"
FT   CDS             22867. .33555
FT                   /*tag= w
FT                   /gene= "megAII"
FT                   /product= "megalomicin 6-deoxyerythronolide B synthase 2"
FT                   /note= "polyketide synthase, encodes AAB82213"
FT   misc_feature    22957. .24237
FT                   /*tag= x
FT                   /gene= "megAII"
FT                   /function= "KS3"
FT   misc_feature    22957. .24237
FT                   /*tag= ai
FT                   /gene= "megAIII"
FT                   /function= "KS5"
FT   misc_feature    24544. .25581
FT                   /*tag= y
FT                   /gene= "megAII"
FT                   /function= "AT3"
FT   misc_feature    24544. .25581
FT                   /*tag= aj
FT                   /gene= "megAIII"
FT                   /function= "AT5"
FT   misc_feature    26230. .26733
FT                   /*tag= z
FT                   /gene= "megAII"
FT                   /function= "KR3 (inactive)"
FT   misc_feature    26230. .26733
FT                   /*tag= ak
FT                   /gene= "megAIII"
FT                   /function= "KR5"
FT   misc_feature    26998. .27258
FT                   /*tag= aa
FT                   /gene= "megAII"
FT                   /function= "ACP3"
FT   misc_feature    26998. .27258
FT                   /*tag= al
FT                   /gene= "megAIII"
FT                   /function= "ACP5"
FT   misc_feature    27393. .28590
FT                   /*tag= ab
FT                   /gene= "megAII"
FT                   /function= "KS4"
FT   misc_feature    27393. .28590
FT                   /*tag= am
FT                   /gene= "megAIII"
FT                   /function= "KS6"
FT   misc_feature    28897. .29931
FT                   /*tag= ac
FT                   /gene= "megAII"
FT                   /function= "AT4"
FT   misc_feature    28897. .29931
FT                   /*tag= an
FT                   /gene= "megAIII"
FT                   /function= "AT6"
FT   misc_feature    29953. .30477
FT                   /*tag= ad
FT                   /gene= "megAII"
FT                   /function= "DH4"
FT   misc_feature    29953. .30477
FT                   /*tag= ao
FT                   /gene= "megAIII"
FT                   /function= "KR6"
FT   misc_feature    31396. .32244
FT                   /*tag= ae
FT                   /gene= "megAII"
FT                   /function= "ER4"
FT   misc_feature    31396. .32244
FT                   /*tag= ap
FT                   /gene= "megAIII"
FT                   /function= "ACP6"
FT   misc_feature    32257. .32799
FT                   /*tag= af
FT                   /gene= "megAII"
FT                   /function= "KR4"
FT   misc_feature    32257. .32799
FT                   /*tag= aq
FT                   /gene= "megAIII"
FT                   /function= "TE"
FT   misc_feature    33052. .33312
FT                   /*tag= ag
FT                   /gene= "megAII"
FT                   /function= "ACP4"
FT   CDS             33666. .43271
FT                   /*tag= ah
FT                   /gene= "megAIII"
FT                   /product= "megalomicin 6-deoxyerythronolide B synthase 3"

  Query Match             15.5%;  Score 192.8;  DB 4;  Length 47981;
  Best Local Similarity   53.9%;  Pred. No. 4.7e-22;
  Matches  526;  Conservative    0;  Mismatches  417;  Indels   33;  Gaps    5;

Qy       149 TGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCGGCCCGCGACG 208
```

```
Db       3637 TGGGGGTGTCGTTCTTCTTTATTCTCAGCGGTTTCGTGCTGACCTGGTCGGCGCGGGCCA 3696

Qy        209 AGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCGCTGCATCTCG 268

Db       3697 GCGACTCGGTGTGGTCGTTCTGGCGCAGACGGGTCTGCAAGCTCTTCCCCAACCACCTGG 3756

Qy        269 CGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCGGTACTTCCCGGCG 328

Db       3757 TCAC---CGCCTTCGCCGCCGTGGTGTTGTTCCTGGTCACCGGGCAGGCGGTGAGCG--- 3810

Qy        329 GTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCCTGGCTTCCCGACG 388

Db       3811 ---------GTGAGGCGCTGATCCCGAACCTCCTGCTGATCCACGCCTGGTTCCCGGCCC 3861

Qy        389 CGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGTGAGATCGCCTTCT 448

Db       3862 TGGAGATCTCCTTCGGCATCAACCCGGTGAGCTGGTCGTTGGCCTGCGAGGCGTTCTTCT 3921

Qy        449 ATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGGCGGCTGTGGT 508

Db       3922 ACCTGTGCTTCCCGCTGTTCCTGTTCTGGATCTCCGGTATCCGCCCGGAGCGGCTGTGGG 3981

Qy        509 GGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCC 568

Db       3982 CCTGGGCCGCCGTGGTGTTCGCCGCGATCTGGGCGGTACCGGTGGTCGCCGACCTCCTGC 4041

Qy        569 CGGCGAG---------CGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTCG 619

Db       4042 TGCCGAGTTCCCCGCCGCTGATCCCGGGGCTTGAGTACTCCGCCATCCAGGACTGGTTCC 4101

Qy        620 CGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTGA 679

Db       4102 TCTACACCTTCCCTGCGACGCGGAGCCTGGAGTTCATCCTCGGGATCATCCTGGCCCGCA 4161

Qy        680 TCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCG 739

Db       4162 TCCTGATCACCGGTCGGTGGATCAACGTCGGGCTGCTCCCCGCGGTGCTGTTGTTCCCGG 4221

Qy        740 CGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCG 799

Db       4222 TCTTCTTCGTCGCCTCGCTCTTCCTGCCGGGTGTCTACGCCATCTCCTCGTCGATGATGA 4281

Qy        800 TACCGGCCGCGCTGCTGATCACGGCGCTGGCCGAC---GCCGACGTGCACGGCCGGCGCA 856

Db       4282 TCCTTCCCCTGGTTCTGATCATCGCCAGCGGCGCGACGGCCGACCTCCAGCAGAAGCGCA 4341

Qy        857 CGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGG 916

Db       4342 CCTTCATGCGTAACCGGGTGATGGTGTGGCTCGGCGACGTCTCCTTCGCGCTCTACATGG 4401

Qy        917 TCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGC 976

Db       4402 TCCACTTCCTGGTGATCGTCTACGGGGCGGACCTGCTGGGGTTCAGCCAGACCGAGGACG 4461

Qy        977 GGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCC 1036

Db       4462 CCCCGCTGGG------TCTCGCACTCTTCATGATCATTCCGTTCCTCGCGGTCTCCCTGG 4515

Qy       1037 TGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCC 1096

Db       4516 TGCTGTCGTGGCTGCTGTACAGGTTCGTCGAGCTACCCGTCATGCGTAACTGGGCCCGCC 4575

Qy       1097 GCAGGTCCGCCTCCCG 1112

Db       4576 CGGCCTCCGCCCGGCG 4591


Search completed: August 25, 2006, 22:07:41
Job time : 917 secs
```

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2006, 21:50:58 ; Search time 6686 Seconds
(without alignments)
10412.738 Million cell updates/sec

Title: US-10-680-860A-13
Perfect score: 1245
Sequence: 1 gtgctccagcgcgtcgatct.....aacccaccccgagaggatga 1245

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est3:*
3: gb_est4:*
4: gb_est5:*
5: gb_est6:*
6: gb_htc:*
7: gb_est2:*
8: gb_est7:*
9: gb_est8:*
10: gb_est9:*
11: gb_gss1:*
12: gb_gss2:*
13: gb_gss3:*
14: gb_gss4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

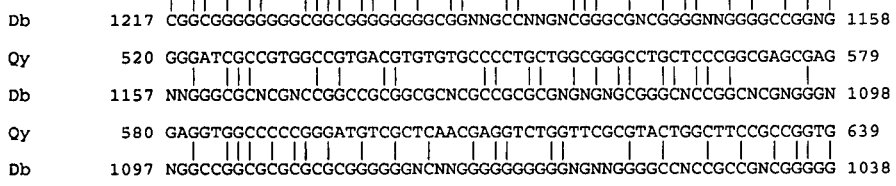
SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	90.6	7.3	1350	2	BM450237	BM450237 AGENCOURT
c	2	88.4	7.1	2243	14	AG381986	AG381986 Mus muscu
	3	84.6	6.8	1798	14	AG171124	AG171124 Pan trogl
	4	84.4	6.8	932	14	CNS0072Q	AL066742 Drosophil
c	5	83.8	6.7	1878	14	AG393083	AG393083 Mus muscu
c	6	81.4	6.5	1625	14	AG043477	AG043477 Pan trogl
	7	80.6	6.5	982	3	BQ687717	BQ687717 AGENCOURT
c	8	80.6	6.5	1428	10	DV779809	DV779809 Hw_Fat_35
c	9	80.4	6.5	1137	2	BG809979	BG809979 mgct002xd
	10	80	6.4	1516	2	BG809984	BG809984 mgct002xd
	11	79.8	6.4	935	14	CNS006XK	AL066051 Drosophil
	12	79.6	6.4	1610	12	BZ569386	BZ569386 pacs2-164
c	13	78.8	6.3	989	14	AG063827	AG063827 Pan trogl
	14	78.6	6.3	1015	12	BZ569259	BZ569259 pacs2-164
c	15	78	6.3	1370	14	AG072551	AG072551 Pan trogl
c	16	77.8	6.2	1598	14	AG030579	AG030579 Pan trogl
	17	77.4	6.2	925	14	CNS0091P	AL053013 Drosophil
c	18	77.4	6.2	932	14	CNS0072Q	AL066742 Drosophil
c	19	76.8	6.2	1569	14	AG341503	AG341503 Mus muscu
	20	76.8	6.2	1956	12	CG754548	CG754548 P050-1-A1
	21	76.4	6.1	1392	14	AG390642	AG390642 Mus muscu
c	22	76	6.1	925	14	CNS0091P	AL053013 Drosophil
c	23	76	6.1	986	14	AG074680	AG074680 Pan trogl
	24	76	6.1	1332	1	AM105094	AM105094 AM105094
	25	76	6.1	1538	14	AG030607	AG030607 Pan trogl
c	26	75.8	6.1	832	14	AG108460	AG108460 Pan trogl
c	27	75.8	6.1	1165	14	AG030649	AG030649 Pan trogl
c	28	74.8	6.0	1332	1	AM105094	AM105094 AM105094
	29	74.4	6.0	1651	2	BG809816	BG809816 mgct001xk
c	30	74.2	6.0	1169	3	BQ922913	BQ922913 AGENCOURT
	31	73.8	5.9	888	2	BG809572	BG809572 mgct001xa
c	32	73.6	5.9	1088	14	AG075424	AG075424 Pan trogl
	33	73.6	5.9	1421	12	BZ569488	BZ569488 pacs2-164
	34	73.6	5.9	1780	14	AG448243	AG448243 Mus muscu
	35	73.4	5.9	1084	14	AG058435	AG058435 Pan trogl
c	36	73.4	5.9	1348	12	CG752544	CG752544 P047-2-C1
	37	73.2	5.9	1131	14	AG042920	AG042920 Pan trogl
c	38	73.2	5.9	1278	14	AG060116	AG060116 Pan trogl
c	39	73.2	5.9	1366	14	AG065058	AG065058 Pan trogl
	40	73.2	5.9	1542	14	AG032943	AG032943 Pan trogl
c	41	73.2	5.9	1800	13	CL494055	CL494055 SAIL_589_
	42	73	5.9	1189	14	AG030608	AG030608 Pan trogl
	43	73	5.9	1299	14	AG039481	AG039481 Pan trogl
	44	73	5.9	1359	13	CL507830	CL507830 SAIL_788_
c	45	73	5.9	1536	13	CL507803	CL507803 SAIL_788_

ALIGNMENTS

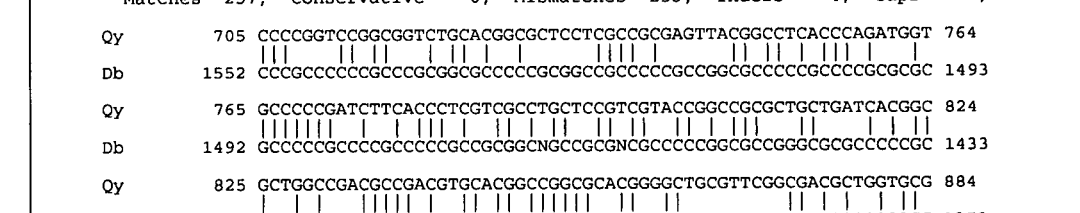
RESULT 1
BM450237/c
LOCUS BM450237 1350 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6393396 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528315 5', mRNA sequence.
ACCESSION BM450237
VERSION BM450237.1 GI:18499277
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1350)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12204 row: p column: 12
High quality sequence stop: 370.
FEATURES Location/Qualifiers
source 1. .1350
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5528315"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."
ORIGIN

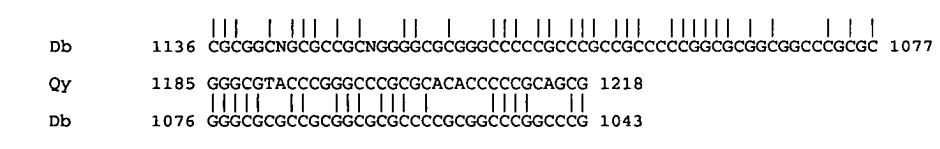
Query Match 7.3%; Score 90.6; DB 2; Length 1350;
Best Local Similarity 42.0%; Pred. No. 7.2e-08;
Matches 320; Conservative 0; Mismatches 441; Indels 1; Gaps 1;



RESULT 2
AG381986/c
LOCUS AG381986 2243 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-192D21.TJ, genomic survey sequence.
ACCESSION AG381986
VERSION AG381986.1 GI:47993191
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 2243)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .2243
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-192D21.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 7.1%; Score 88.4; DB 14; Length 2243;
Best Local Similarity 50.0%; Pred. No. 2e-07;
Matches 257; Conservative 0; Mismatches 253; Indels 4; Gaps 2;

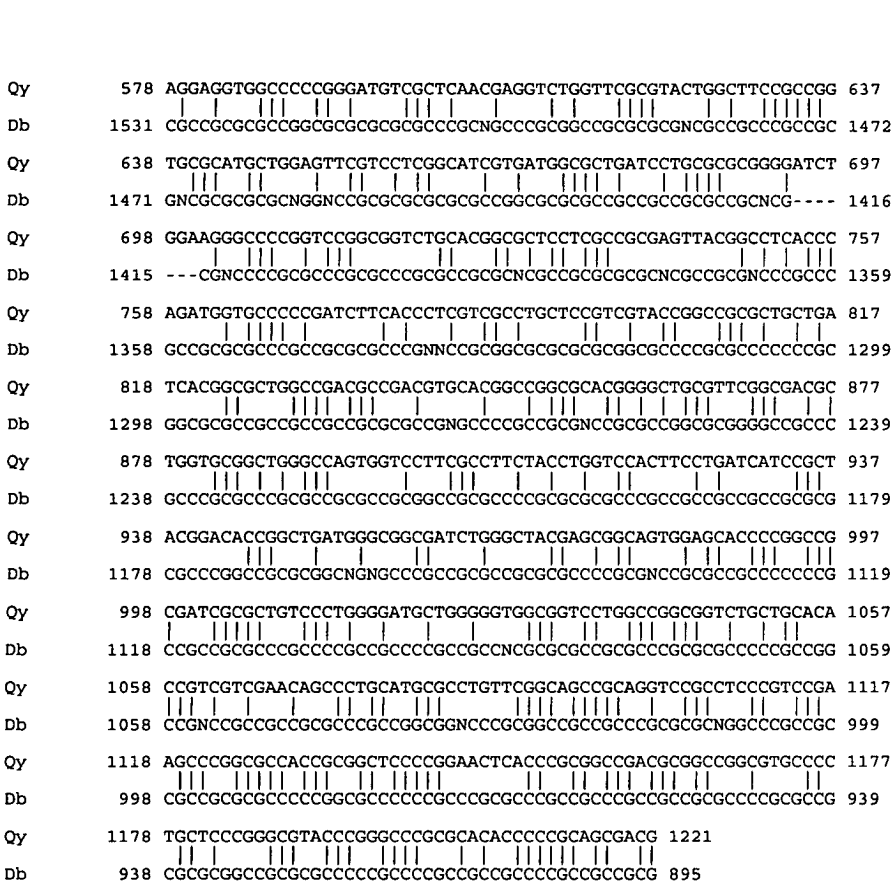




```
Db       1136 CGCGGCNGCGCCGCNGGGGCGCGGGCCCCCGCCCGCCGCCCCCGGCGCGGCGGCCCGCGC 1077

Qy       1185 GGGCGTACCCGGGCCCGCGCACACCCCCGCAGCG 1218
              ||||| || ||| ||| |     ||||  ||
Db       1076 GGGCGCGCCGCGGCGCGCCCCGCGGCCCGGCCCG 1043


RESULT 3
AG171124
LOCUS       AG171124               1798 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
            sequence.
ACCESSION   AG171124
VERSION     AG171124.1  GI:16700802
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                Sequencing: TJ
             LIBRARY
                Vector    : pBACe3.6
                R.Site 1  : EcoRI
                R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1798
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match             6.8%;  Score 84.6;  DB 14;  Length 1798;
  Best Local Similarity   34.3%;  Pred. No. 1.2e-06;
  Matches  422;  Conservative    0;  Mismatches  796;  Indels   14;  Gaps    3;

Qy          4 CTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTA 63
              |                                  || |  | ||       | |   |||
Db          4 CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCCNNCCGGGGNNNNCCCCNNCCGGNN 63

Qy         64 TTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTG 123
                    |                                |                |
Db         64 NNNNNNNNCCNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNCNCNNNNCNNNNNN 123

Qy        124 CTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTC 183
                   |
Db        124 NNNNNNNNNNCNNNNNNNCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCNNNN 183

Qy        184 GTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATC 243
                  |||  |  ||  |  ||    ||    || || |        |||||  ||   |

Db        184 CCGGCCGCGCGCGCGCGCNCCGCGCGGGCGCGCGCGCGGCGCGCGCNGCGGCCGCGGCGC 243

Qy        244 GCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCC 303
                  |||  |      | || ||      |  ||    |
Db        244 GCGCGGCCGCCGCGGCGGCCCNNNNCCGCGGGGCGGNNNCNNNNNNNNNNNNNNNNNNN 303

Qy        304 CTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTG 363
                                              |||      |     |       ||      |
Db        304 NNNCNCCNNNNNNNNNNNNNNNNNNCCGCCGCCNCCGCNCCGGCCCCCGCCCGCCGCCGC 363

Qy        364 CTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGG 423
                | ||      | | |  ||           | | | |||  |   | | ||||  || ||
Db        364 CGCGGCGCCGGCCGCCCGCCNCCGCNNCGGCNCCCCGCCGGCCGCGCCNCGCCGCGCCGG 423

Qy        424 TCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGC 483
                 ||| | ||       ||||      |  |  | | | |  |    |   || |
Db        424 CCGCCCCCCGCGCNCGCCGCCCCGCCGCGCCCCCCCGCCCCCGGNNGGGGNNNGGGGGNG 483

Qy        484 AGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGT 543
                ||      ||      |||        || |||| | || |  ||  |      | | |
Db        484 GGGGGNGGCGNNNNNCGGNNNNNNGGGGGCGCGGGGGNGGGCGGGGGNNNNGNGNGGNNN 543

Qy        544 GTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTC 603
                |       || ||  |||  |     || |   | |   | |       |||  |
Db        544 GGNGGGNCGCGGGGNGGCGCGNNNGGGGGGGNGGGGCGNGCGNGGGGGGGGGGGGNNNGG 603

Qy        604 AACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGC 663
                  | | |  | || |  | ||  |  |||  | ||  |   |   ||  | ||||
Db        604 GGNGGCGGCGCGCGCGGGCGCGCGCGGCGCGCGGGGNGCGCGGGCG---CCGCGCGCGGC 660

Qy        664 ATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGC 723
                  |  |||   |  |  |||| | |  | |  |||  |  |  |  |  |  |  |
Db        661 GCGGGCGCGGCCGCGGGCGCCCGCGGCGCGCGCGGNCGNGGCGCGCGCGCGCGCGCGC 720

Qy        724 ACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTC 783
                ||      | ||| ||| |   ||  | | | |      |  | | ||  |   | | |
Db        721 GCGCGCGCGCGCGCGGCGCGGCGCGCGCGCGCGCGCGNCGCGCNCGCGCGCGCCGGNCGCCC 780

Qy        784 GTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTG 843
                | |||     |||     |  ||  ||| || | |  |  ||||  ||||    ||    |
Db        781 GCCGCGGCGCCCGCNCNCCGCGCGCCGCCGCGGCGCNNNGCGC--GCCGCGCNCGCGCCG 838

Qy        844 CACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTC 903
                | || |   | | ||| ||      | ||| ||| |      ||  | |  |  |  |
Db        839 CGCGNCGCCCCCGCGGCGCGCNCGCCCGCGNCGCCGCGCGCGCNCGCCGCGCCGCGCGCC 898

Qy        904 GCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGAT 963
                |   |  |  |  |       |    | ||||  ||  | |  |  |      || ||
Db        899 NCGCGCCCGCGCGNNCCGCGGCGCCGCGCCCGCCNCGCCCGCNCGGCGCNCNNCGCCGCC 958

Qy        964 CTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTG 1023
                | | ||  |  || || | || || | ||| |||||  |    |    ||   ||
Db        959 CCGCGCCNCNCGCCGCCGCGGCGCNCGCCGCCCGCGCGCCGCCGCCGCCCNCGGCGC--- 1015

Qy       1024 GGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGC 1083
                      ||    |    |  ||  |       || || ||  |    |    ||||
Db       1016 ------GCCNGCNGCNCGCCCGCGCGCCNNNNNNGCGCCGGCGCGCGCGCGCGNGCGCGC 1069

Qy       1084 CTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGG 1143
                | |   ||  | |  | | | || |    |||||    | |  | ||   |||
Db       1070 GCGCGCNNNCGCGCCCGCGNCNCNCCGCGNCGCCCGCCCGCGCGCGCGNCCGCGGGCCGC 1129

Qy       1144 AACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGCG 1203
                | | ||||| ||      ||  | |||     | || | || |||   | | || |||
Db       1130 CNCNCGCCCGCCGCGCNNCCGCNCNGCGGCGGCGCGCGCGCGCGCGCGNCGGCGCGCGCC 1189

Qy       1204 CACACCCCCGCAGCGACGAACGAACCCACCCC 1235
                | | | ||| ||  ||  ||     | || |
Db       1190 GGCGCGCNCGCGGCNNCGCCCGCCGGCGCCGC 1221


RESULT 4
CNS0072Q
```

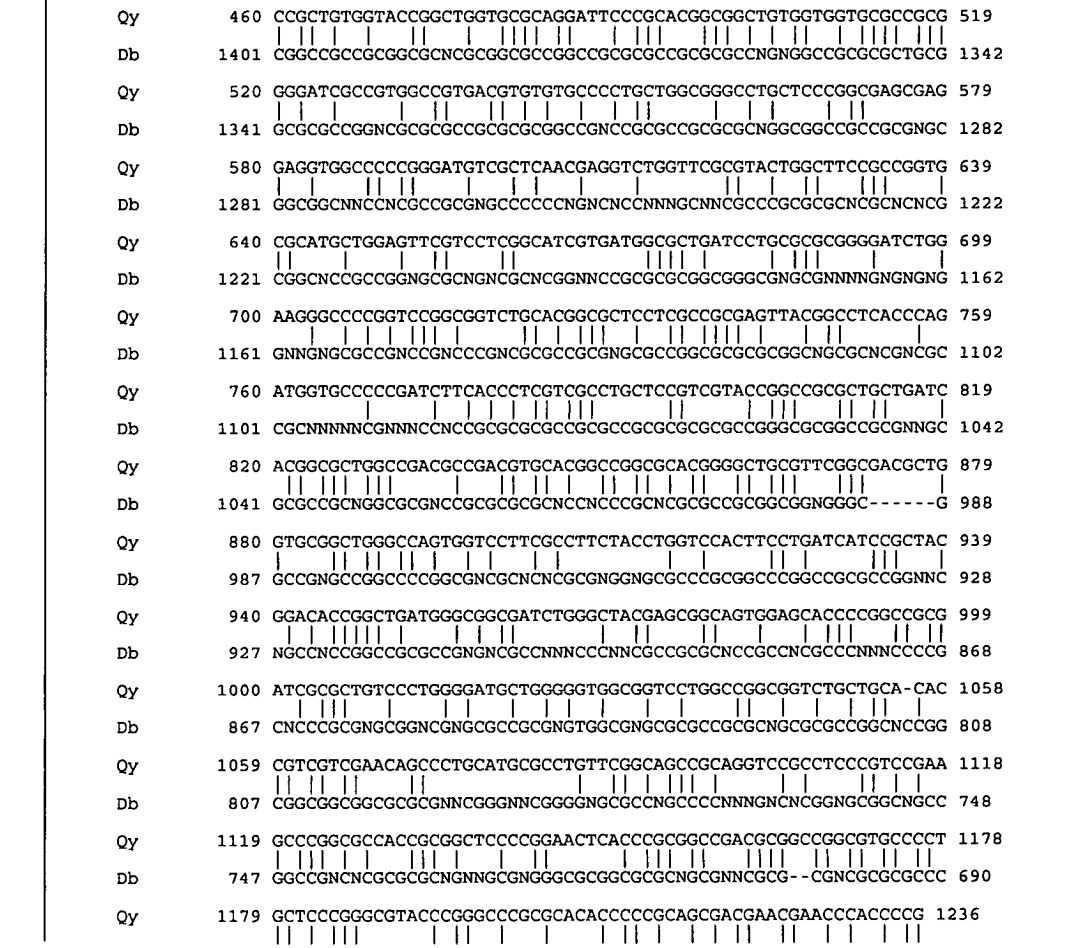
```
LOCUS       CNS0072Q                 932 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066742
VERSION     AL066742.1  GI:4945205
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14B09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             6.8%;  Score 84.4;  DB 14;  Length 932;
  Best Local Similarity   31.2%;  Pred. No. 1.3e-06;
  Matches  148;  Conservative  114;  Mismatches  212;  Indels    0;  Gaps    0;

Qy        418 AGCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTG 477
              || |||:|: | |: ||| | : |  :| |  || ::  || | | | | | |::|| :
Db        363 AGGTGGKCBGTGTBGTGTTTGGKTGTTKTTTGCCTCKSTCTCGCTCCGCGTTYCYBGCGS 422

Qy        478 GTGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTG 537
                :|:|  : :   |:|      ::| :    :  :::  |: :|   :|:   :  ::
Db        423 TCKCKCGCKCKCTGCSCTGTTTYKYTCKCTGCSCTSSBGYCKCKGKCKKCBCGCKGGTKB 482

Qy        538 ACGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATG 597
              | |:|  :  |:|  || ||| : |    :||:  : |  | |: |  |  | ::
Db        483 CKGGGKGGCBKGCKGGCGGGGGGGVCGRGGGSGGSCGSTGGCGGGKGGCGCKGCTSKGGC 542

Qy        598 TCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTC 657
              | :      | ||   |:  :||:: | ||  |  |: || |::: |::|:: :  || :
Db        543 TGBGSTGCGGGGGSKSGSGGSGCSBGCGGGGSTKGGSGGGGGSBSBTKYTBKGKGGCGGS 602

Qy        658 CTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCG 717
              : :  :  :  | ||||: |     ||| |::: |  |  |  :||| |    |::|:|
Db        603 SGSCCSGGSCRGGGGGCGSCGGSGGCGCGGGSSSCGCSCSCGCGSGGCGCGCCCGCSSGSG 662

Qy        718 GTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTC 777
              | : |   |: :  |  |::||| |   |  |:: | | ||| |: | :||   :
Db        663 GCSSGGSGGSGSGCCGGCSSCGCCSGCCCCSSGCSCCCCCGCGCGSCGGSCCGCGCSCSG 722

Qy        778 ACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCC 837
                |::   | |: :| ||: |   :|| ::::|| |: |  : ||:||| :   : ||:|
```

```
Db        723 GCSSCGCSCCCSCSCCCCSCCCCSSCGCSSSSGCGGSGGCSSGCGSCGCGSCGGSGCGSC 782

Qy        838 GACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGC 891
              | || :| ::: ||: :: :| |:: : :   ||:| ::  :: |::|| ::|:
Db        783 GGCGGSCGSSSGCGSGSSGSGCGSSSSGSSCGGGSGCSSGCSSGGSSGCCSSGS 836


RESULT 5
AG393083/c
LOCUS       AG393083                 1878 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-211E23.TJ, genomic survey
            sequence.
ACCESSION   AG393083
VERSION     AG393083.1  GI:48018931
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1878)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1878
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-211E23.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             6.7%;  Score 83.8;  DB 14;  Length 1878;
  Best Local Similarity   45.0%;  Pred. No. 1.7e-06;
  Matches  344;  Conservative    0;  Mismatches  412;  Indels    8;  Gaps    2;

Qy        459 CCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGC 518
              | |||    |  |||   |  |||| |   ||  | | ||| | | |   | ||||||||
Db       1651 CGCGCCCGCGCGCCGCGCGCCGCGCGCGCCCCCGCCGCCGCGCCCGCGCGCGCGCCGC 1592

Qy        519 GGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAG-CG 577
              |    |  | | || |  | | ||| |   |   | ||  | | | ||| | ||
Db       1591 GNCCCGCCGCNCCGNCGCGCGCGGCGCGCCGCGCGCGCGCCCGCCGCGCNGCGCGCCG 1532
```



```
RESULT 6
AG043477/c
LOCUS       AG043477               1625 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-021M21.F, genomic survey sequence.
ACCESSION   AG043477
VERSION     AG043477.1  GI:16572202
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1625)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1625
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-021M21.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
```

Query Match 6.5%; Score 81.4; DB 14; Length 1625;
Best Local Similarity 40.6%; Pred. No. 5.1e-06;
Matches 316; Conservative 0; Mismatches 453; Indels 9; Gaps 3;



```
Db        689 GCGCGCGGCNCGNGCNGGNGCGCGGNNGGCGCCGGNNGGGGCGGGCGNNCGCGGCGCG 632


RESULT 7
BQ687717
LOCUS       BQ687717                982 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8343763 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6248424
            5', mRNA sequence.
ACCESSION   BQ687717
VERSION     BQ687717.1  GI:21813033
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2388  row: e  column: 01
            High quality sequence stop: 216.
FEATURES             Location/Qualifiers
     source          1. .982
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6248424"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of  California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             6.5%;  Score 80.6;  DB 3;  Length 982;
  Best Local Similarity   44.7%;  Pred. No. 7.3e-06;
  Matches  332;  Conservative    0;  Mismatches  407;  Indels    4;  Gaps    2;

Qy        472 CGGCTGGTGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTG 531
              |||  || |  | |      |      ||      | ||      | ||||   |||
Db        201 CGGAGGGGGGNCGGNGNNGNGGNGGNNGGGNNNNNGGGGNNGNGGGGGGGCCGGCCCGC 260

Qy        532 GCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCC 591
              | || |     ||     ||||| || |     |  | ||   | |  ||  |
Db        261 CGGCCGCCGCGCCCCGCCGCNGGGGCGGCCCCGNGGGGGCNGGGCCCCGCGCCGGGGCGG 320

Qy        592 GGGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAG 651
              |||    ||| |  ||  |   ||  ||||  |       |||   | || |   |  |
Db        321 GGGGCCGCGCCCCCCGGCGGGCGGGGCGCGGGCGCCGCGCCGGGCGCGCCCCCCCCCGCC 380

Qy        652 TTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGT 711
                || ||   ||  ||     | |  |      || ||          |  |   ||||  |
Db        381 CCCGCCCCGCGCCGCGCCCCCGGGGCGGCGGCGCCCGGCCCCCGCCCGCGCGGCCGGGCG 440

Qy        712 CCGGCGGTCTGCACGGCGCTCCTCGC---CGCGAGTTACGGCCTCACCCAGATGGTGCCC 768
              |||||  | |  | |||  || | |    |    |   |||||  | ||  |   |  | |
Db        441 GCGGCGCCCCGGGCCGCGGGCCGCCCGCGCCGCCGCCGCGGCCGCCCCGGGCGCGGCCGC 500

Qy        769 CCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTG 828
                      ||  ||     | |   || ||  |   ||||| |  |     |  |||
Db        501 GGCGGGCCGCCCGNCGCGCGCCGGGGCGCCGCCCGCCCCGCGCGGGGGCGGCCCCCGCCC 560

Qy        829 GCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTG 888
              |||  | | | || ||   ||||  || |||||  |||  ||||| |||  |  |||| |
Db        561 GCCCNCCCGGGCGCGCCGCGCCGCGGCGCGGGGGCGCGGCCGGCGCCGCCCGCCCGGCCG 620

Qy        889 GGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGG 948
              |||  | ||  |    |   |  |  | | |  |    |  |  | | |          ||||
Db        621 CGCCGGCGGGGCCGGCCGGCCGCGCGCGGCGCCGCGCCCGCCGCCGGGGCGCCGGCCCGG 680

Qy        949 CTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTG 1008
                  ||| || |   |||   || || |    |     | | | || ||   ||||
Db        681 GGCGGGGGGGGGGCGGGGGGGCGCGCGCCGGGCGGCCCGGCGCGGCCCCCGCGGGCGCGC 740

Qy        1009 TCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAA 1068
               | | ||   | | ||| ||  |    | |  ||||   || || |   |    ||
Db         741 GGGCCGCGGGCNCCGCGGGGGGGCGCGGGGCCGCGCGGGGCGCGGCCCCNGGGGCCCGGC 800

Qy        1069 CAGCCCTGCATGCGCCTGTT-CGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGC 1127
               | | || ||||  |   || | |||     | ||| | | ||| |    |||| |||
Db         801 CGCCGCGGCGCGCGCGCCGCGCCCGCCCNCCGCCCGNCGCNCGNCGCCGGCCCCCCGC 860

Qy        1128 CACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGG 1187
               |  ||| |    || |  | | ||||    |   || ||| |    | ||  | | |||
Db         861 CCGCGCCGGCGGCCCGCGCCCGCCCGGCCGCCGGGCCGCCCGGCCCCGCCCNCGCGCGGC 920

Qy        1188 CGTACCCGGGCCCGCGCACACCC 1210
                ||   |  ||| ||||| | |||
Db         921 CGCCGCGTGGCGCGCGCGCCC 943


RESULT 8
DV779809/c
LOCUS       DV779809                1428 bp    mRNA    linear   EST 23-NOV-2005
DEFINITION  Hw_Fat_35_050805_D06 Bos taurus CF-24-HW fat cDNA library Bos
            taurus cDNA, mRNA sequence.
ACCESSION   DV779809
VERSION     DV779809.1  GI:82632685
KEYWORDS    EST.
SOURCE      Bos taurus (cattle)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1428)
  AUTHORS   Yoon,D.H., Lee,S.H., Park,E.W., Cho,Y.M., Lee,J.H., Kim,H.,
            Kim,H.Y., Park,J.H. and Oh,S.J.
  TITLE     Gene Expression Profiling of the Bovine liver, adipose, and
            skeletal muscle
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Dr. Du-Hak Yoon
            National Livestock Research Institute, RDA
            564 Omockchun-dong, Suwon, 441-350, Korea
            Tel: 82 31 290 1593
            Fax: 82 31 290 1792
            Email: dhyoon@rda.go.kr.
FEATURES             Location/Qualifiers
     source          1. .1428
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /sex="Four males mixed"
                     /cell_type="adipocyte"
                     /dev_stage="24 months old"
                     /lab_host="XL1-BlueMRF' strain"
                     /clone_lib="Bos taurus CF-24-HW fat cDNA library"
                     /note="Organ: adipose tissue; Vector: Uni-ZAPXR; Site_1:
                     EcoRI; Site_2: Xho I"
ORIGIN
```

```
  Query Match             6.5%;  Score 80.6;  DB 10;  Length 1428;
  Best Local Similarity   48.2%;  Pred. No. 7.3e-06;
  Matches  275;  Conservative    0;  Mismatches  292;  Indels    4;  Gaps    2;

Qy        667 GTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACG 726
              | |  |||||    || | ||| | |   |  | |  ||||||   ||    |    |
Db       1088 GCGGCGGCGCGCCGCCGGGGCGGGCGCCGCGCGGCGCCCCCCGGCGGGGGCCGCGCGCCC 1029

Qy        727 GCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTC 786
              ||       |  ||||      | |  |  || |   |  |||||  |  | ||| || |
Db       1028 GCCGCGGCCCGCGCGCCCGGGGCCGCCGGCCCGCGCGCCGCCCCGCGCGCCCCCCCCGGC 969

Qy        787 GCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCAC 846
              ||| || ||      |||  ||| |  |     |  || ||| |   ||  |  | |
Db        968 GCCCGCGCCCCGCCCCCGCGCGCCCCCCGCGCCGNNGGGCCGGCGGGGCCCCGCCGGGGC 909

Qy        847 GGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCC 906
              ||| ||||| ||  || |    | |  | |  | || || | ||| |  | ||  || |
Db        908 GGCGGGCGCCCGCCGCCGGCCGCCCCCGCNCCCGCGCCGCCGCGCCCGGCGGCCGCCGGC 849

Qy        907 TTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATG--GGCGGCGATC 964
              ||    || |  |     |  || |  |   |  || | |  |  | || |    |
Db        848 GCGGCCCCCCGCCCCGCCGCCGCCCCCCCCCCCCCCCGGCGCCGGCCGGCGCCGCCCGCC 789

Qy        965 TGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGG 1024
              ||||   |  | ||    ||  | ||| ||||    ||||  |||    |     ||
Db        788 CGGGCCGGGCCCCGCCCGCGCGGCCGCCGCCCGCCCCGGCGCCCGCCCCCCCGCCCGCGG 729

Qy       1025 GGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCC 1084
                |  || | || |  | |||     |    | ||  |  ||  |    | |     ||
Db        728 CCGGCGCCGGCCGGNGCCGCGCGGCCCGCNGCCCCCCCCCCGCCCCCNGCCGGCGCGGCG 669

Qy       1085 TGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGA 1144
                |  | ||  |||| || |  |  | |  |||  ||||| | |  | || || ||||||
Db        668 CGCCCCGCGCCCGCCGGCCGCCGCCGCCGCCGCCGCCCGNCCCGCCGGCCGCCCCCCGG- 610

Qy       1145 ACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGCGC 1204
                |   || ||| ||  | ||  | ||    |  | ||  |||  |   |||||||||  | |
Db        609 -CCGCCCGGCGCCCNCCCCGCGCCGCCCCCGGCCGCCGCCGCCCCCCCCCGGGCCCCCCC 551

Qy       1205 ACACCCCCGCAGCGACGAACGAACCCACCCC 1235
                |||||||  ||| |   |   ||| | ||
Db        550 CGGCCCCCGCCGCGCCCCCCCCCCCCCCGCC 520


RESULT 9
BG809979/c
LOCUS       BG809979               1137 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  mgct002xd01f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
            grisea cDNA clone mgct002xd01f 5', mRNA sequence.
ACCESSION   BG809979
VERSION     BG809979.1  GI:14180959
KEYWORDS    EST.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 1137)
  AUTHORS   Choi,W. and Dean,R.A.
  TITLE     Construction and sequence analysis of an appressorium stage cDNA
            library in the rice blast fungus, Magnaporthe grisea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG).
FEATURES             Location/Qualifiers
     source          1. .1137
                     /organism="Magnaporthe grisea"
                     /mol_type="mRNA"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgct002xd01f"
                     /dev_stage="Germinated conidia on apprressorium-inductive
                     surface"
                     /clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
                     /note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI;
                     Site_2: XhoI; The appressorium formation-specific cDNA
                     library was constructed from conidia germinated for 5-8 hr
                     on an inductive surface. The library contains over 55,000
                     clones with average insert size of 1.5 kbp."
ORIGIN

  Query Match             6.5%;  Score 80.4;  DB 2;  Length 1137;
  Best Local Similarity   44.4%;  Pred. No. 8e-06;
  Matches  353;  Conservative    0;  Mismatches  432;  Indels   10;  Gaps    1;

Qy        444 CTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGGCGGCT 503
              |||    | | ||||||   | | |    | ||  |    |        |||| || || |
Db       1131 CTTTGTTTTTTCGTTCTTCCCGCGCGCGCCGCCCGCCGCCCCCGCGGCCGCGCGCCGCCG 1072

Qy        504 GTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCT 563
              |    |  ||||||      ||||   ||  |  || |    |||| |   || |  |
Db       1071 GCCCCGCGCCGCCGCCTCCGCCGCCCCCGCGCTTCCGCGCCGCCCCCCGGGCGCCGCGCC 1012

Qy        564 GCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTCGCGTA 623
              |   | ||  ||| ||   | | |||| |    |      | |     |   ||
Db       1011 CCGGGCTGCCGGCGGGGGCCCGCCGCCCGCGGCCCCCGCGCGGGCGCGGGCGCCGCCGCG 952

Qy        624 CTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCT 683
              | | |   ||  || |  |    |  |  |  | ||  |   | |  |     |
Db        951 CCGCCCCCGCGCGCGGCCCCCGCGGCGCGCCCGCGNCGCCGCGGGGCGCCCGCCGCGCCG 892

Qy        684 GCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAG 743
              | |||| ||  |          ||| | | |   |  | | || | || |||| || |
Db        891 CCCCGCGCGGCCCCCCCGCCCGGGCGGGCGGCCCCGCGNCCCCGCCCGCCCCGCCCCGGG 832

Qy        744 TTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACC 803
                | ||| |   | |  |   || |  |  | |   ||   | ||| ||  ||
Db        831 CCGCCGCCCCCGGCCGGCCGCCGCCGCCCCCCCCCCGGCCGCGCCGCCCGCCGCCGCCCC 772

Qy        804 GGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCT 863
              |  ||| |  |    | ||| || ||  | ||| | ||  | || || ||
Db        771 GCGCGCCCCCCGCGGCGCGGGGCCGGGGGCCGCGGNCGGCCCCGCCCCGC---------- 722

Qy        864 GCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTT 923
              |||  || || |   |  |||| | ||  | |  ||  || |     || |    |
Db        721 GCGCCCGCCGCCCGCGCGGCGGCCGCGCGGGCGCCCCCGCGGCGCGCCCCCCGCGGCCCG 662

Qy        924 CCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTG 983
              |       |     | |       |||||||      |    | | |||
Db        661 CGGCGCGGCGCCGCNGCGCCGGGCCGGCNGGGGCGGCCCCNGCGNGGGGGCGGGGCGCGC 602

Qy        984 GAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGG 1043
              | || | |  || ||| |||       |   |||  |  | |||  || |  | | || |
Db        601 GCGCCCGCGCGCGGCGCCCGCCGGCCGCGGCGGGGCGGCGCGGGCCGCCGCGCCGCCCCG 542

Qy       1044 CGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTC 1103
              | | | ||    | ||  |         ||   ||  ||||      ||| |||||    |
Db        541 CCGCCCGCCCGCCCCCCCCCGGCGCGCGCGGGGCGGGCGCGGNCCGGGCGGCCGCNCCCC 482

Qy       1104 CGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGC 1163
              |||| |    |       |  | |  ||||  | | |     | | | | | |||  | |
Db        481 CGCCCCGNCCCGCGCCGCGCCCCCGCCCGCCCCCCCGCGCCCCCNCGCGCCCCGCCCCCCC 422

Qy       1164 GGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGCGCACACCCCCGCAGCGACGAA 1223
                ||  |    ||| || |  |||||  ||| | |||| || | || || |   | |
Db        421 CCCCCCCCGCGCCCCGCGCGGGGGCGCCCCCTGCCCCGGGCCCGCCGCCCCGCCCCCCGC 362

Qy       1224 CGAACCCACCCCGAG 1238
              ||  |||  |||| |
Db        361 CGCCCCCCGCCCGGG 347
```

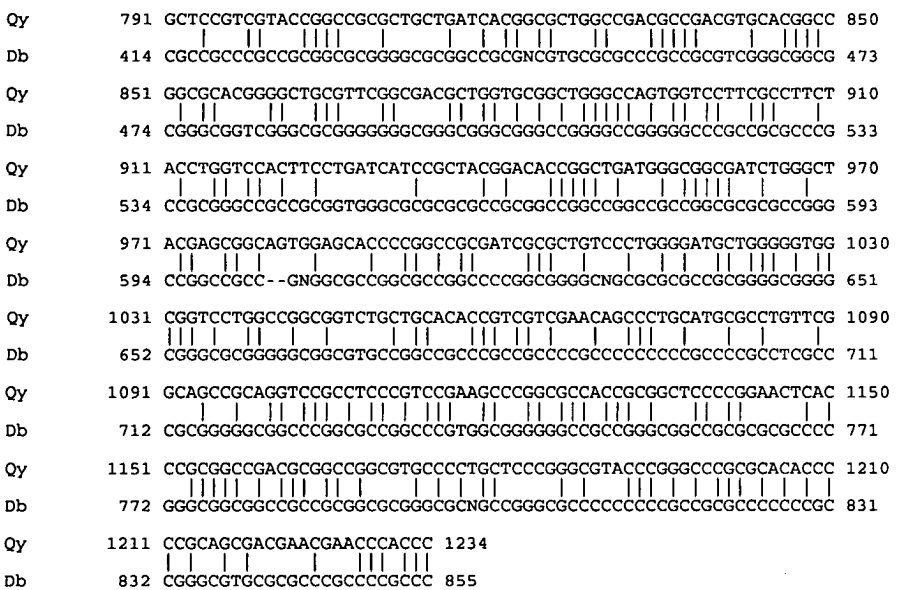
```
RESULT 10
BG809984
LOCUS       BG809984               1516 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  mgct002xd11f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
            grisea cDNA clone mgct002xd11f 5', mRNA sequence.
ACCESSION   BG809984
VERSION     BG809984.1  GI:14180964
KEYWORDS    EST.
SOURCE      Magnaporthe grisea (anamorph: Pyricularia grisea)
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 1516)
  AUTHORS   Choi,W. and Dean,R.A.
  TITLE     Construction and sequence analysis of an appressorium stage cDNA
            library in the rice blast fungus, Magnaporthe grisea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG).
FEATURES             Location/Qualifiers
     source          1. .1516
                     /organism="Magnaporthe grisea"
                     /mol_type="mRNA"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgct002xd11f"
                     /dev_stage="Germinated conidia on apprressorium-inductive
                     surface"
                     /clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
                     /note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI;
                     Site_2: XhoI; The appressorium formation-specific cDNA
                     library was constructed from conidia germinated for 5-8 hr
                     on an inductive surface. The library contains over 55,000
                     clones with average insert size of 1.5 kbp."
ORIGIN

  Query Match             6.4%;  Score 80;  DB 2;  Length 1516;
  Best Local Similarity   43.6%;  Pred. No. 9.7e-06;
  Matches  324;  Conservative    0;  Mismatches  417;  Indels    2;  Gaps    2;

Qy        492 CGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCT 551
              ||  |||||   | |   | ||||     |   ||   ||        | | | | ||
Db        449 CGNGCGGCGCGCGCGCGCGCGCGCGCCCCGCGGGCGCCGGNGNGNGGGGNGCGGCCGC 508

Qy        552 GCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGT 611
              |  || ||| | |   || ||  ||   | || | | ||  ||     |       ||  |
Db        509 GGGGGGGGGGCGGNCNCCNGCNCGCCGCGCGGCGCCGCCGCGGGCNGCCGCGCGCGGCGC 568

Qy        612 CTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGAT 671
                |   || |     |   | |||||     | |  ||         | | |   |
Db        569 GCGCNGCGGGGCGGGCNCGCGGCCGGGCGCCGCGGGGGCCGCGCGGCGCCGGCGCCGCCC 628

Qy        672 GGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCT 731
              ||||  |  |   |||   | |  | |     || |||  || || | |         ||
Db        629 NGCGCCGCCCGCCCGCCNCGCGCCCGGCGCCCGCGCCGCGCCCGCCGGCGCGGGCCGGCG 688

Qy        732 CCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTG 791
                ||       ||||  |   |    || |  | |  |  |  |   |  |   |
Db        689 NGNGCGCGGCGCGGCCGGCGGCGGGCGCGCGGCGGGGCGGGGCGCCGNCGCGGNNGGGGG 748

Qy        792 CTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCG 851
              | ||  ||  | | | |||  || |  | |||||   | |  |  || || || | ||||
Db        749 C-CCCNCGCGCGGCCGGCGGNGCCGGGCGCGGCGGCNGGCCGCCGCGGCGNGCGC-GCCG 806

Qy        852 GCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTA 911
```

```
              |||| ||  || |||    ||| |    | |||   ||| | | || |   || |
Db        807 GCGCGCGCCGCCGCGNNGCGCGCCAGGCGCGCGCGCGGGGCGGGGGGCGGNCGGCGCGCG 866

Qy        912 CCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTA 971
              ||  |    |    |    |    ||        || | | | |   ||     |
Db        867 CCGCGCGGGCNGGGCGCGGGGCGCGCGCGCCGCGGCGCGGCCGCGGCNCGCNGCCGCGGC 926

Qy        972 CGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGC 1031
              ||  | ||   ||  | ||||||||  ||  ||| |   | | || |  || || ||  ||
Db        927 CGCCCCGCGCCGGGCCGCCCCGGCCCCGCGCGCCCCCCCGCCGGCGGCGCGGGCGGCCGC 986

Qy       1032 GGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGG 1091
              ||  | |  | |||      |    | |   |  ||   ||||   | |  |     |
Db        987 GGCGCCGCGCCGCGCGGCCCGCGCCNCGCGCCGCGGCGAGCCGCNCCGGACGCCCGCGGC 1046

Qy       1092 CAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACC 1151
              | |||||  |  |  | |||  ||| ||  || |||||||     ||| | ||         |
Db       1047 CCGCCGCGCGCGCCGCGGCCGCCCGGCGCGCGGCGCCGNGNNGGCGCGCCCCGCGNGCGC 1106

Qy       1152 CGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGCGCACACCCC 1211
              |  |||  |   ||   ||  || |||  | |    |  || ||||     | |
Db       1107 GGGNGCCCGCCGCGCGCNCGCNGCCGCGCGCCCNCGNGCGNGCCCGCGCGCGNNNCNCNC 1166

Qy       1212 CGCAGCGACGAACGAACCCACCC 1234
                |  || || |||  | | |||
Db       1167 GCCGCCGCCGCACGCGCGCGCCC 1189
```

```
RESULT 11
CNS006XK
LOCUS       CNS006XK                 935 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14N09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN
```

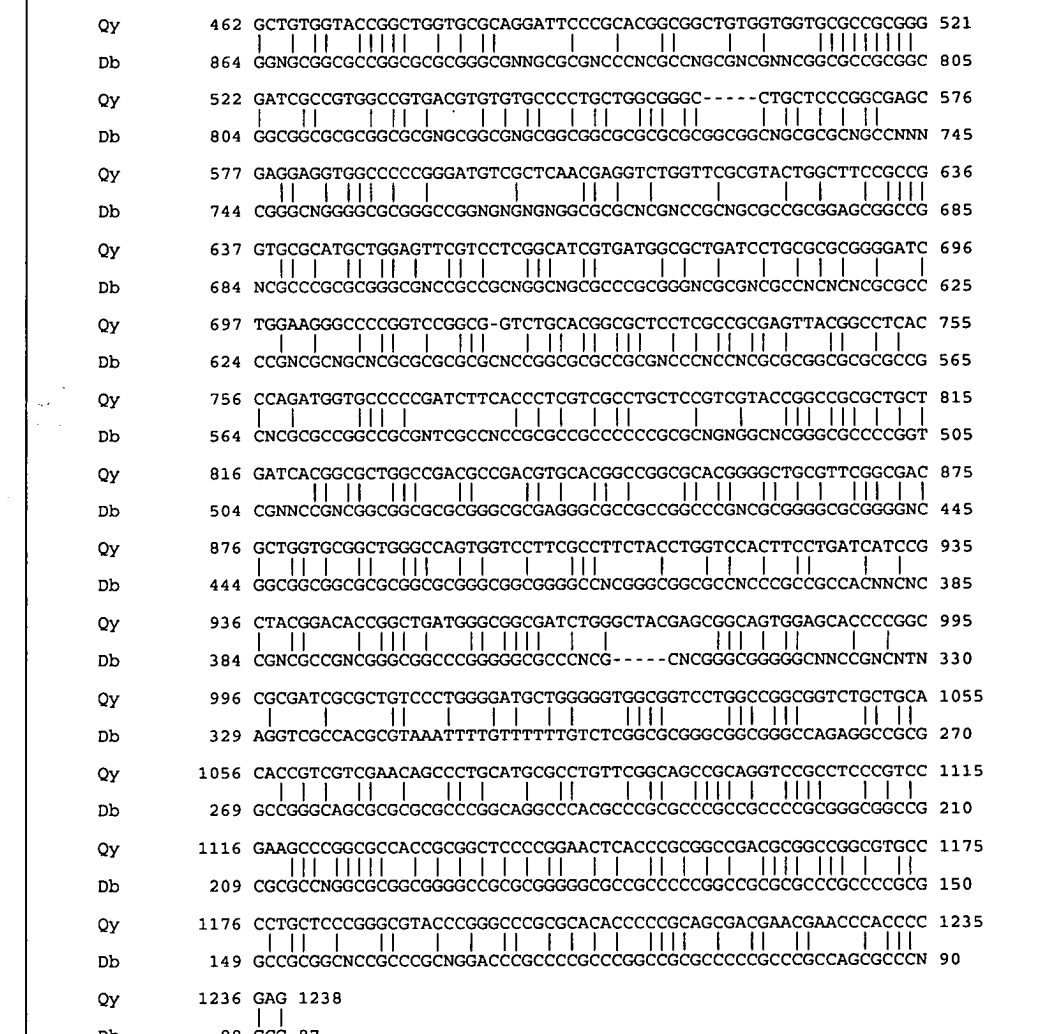
```
  Query Match             6.4%;  Score 79.8;  DB 14;  Length 935;
  Best Local Similarity   30.1%;  Pred. No. 1.1e-05;
  Matches  132;  Conservative  112;  Mismatches  195;  Indels    0;  Gaps    0;

Qy        720 CTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCAC 779
              :|| :|: ::| |  |::|: : | | |: || :::::::  :  | || :   || | :
Db        497 STGTMCDMMVCGCGSCSMCRTSSGSTGCSTCCMYMSSSVSCCSCSGTCCGYCSSCTSCKM 556

Qy        780 CCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGA 839
              |:    |:|  || |:  |:  :| ::::::  :||: |  ::||:  :|| | ||  |
Db        557 CSCTYGCKCGCGCGCSTSCSSSSCCSBBSYSTCSCTBCTKCSSGCSTGSGCTGCCGGGGS 616

Qy        840 CGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTC 899
                || :|  ||  :    |||| ::|  :|:::  :: :: |:::: :||   | |  :
Db        617 GCGGCGSGCGCGCGSGGGCGGGCSSSGGSSGSSSGGSSGSSCGSSSSGSGGGGSGGGCGS 676

Qy        900 CTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGG 959
              ||   :   | :  |:    :  |  |    ||  ||  | :||   | :|::|
Db        677 GGCGGCGCGSCSGCGSCGCSSGCGSGCGSGCGSGGGCGSCGCSCSSCGCGSGCSSGSSGC 736

Qy        960 CGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGAT 1019
              :  | |: :  :: |:  : | |   ::|: |||::  |  :||   : | | :||
Db        737 SCCSCCGSCSCSSSSGSCCSCGCGCGMSMCSGCGGSSCGGCCSGCCGCSSCGCCSGGCCS 796

Qy       1020 GCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCAT 1079
              || || |  :|: | |  ||:::| :| | |: |  | :|   | ||  :  : |
Db        797 GCCGGCGCGSGSCGGCGCGGSSSGGSGSCCGSCGGGCGSCCSGGGCGCSSSCSGCGCGSG 856

Qy       1080 GCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCC 1139
              |::|| :  ||:  |: :: || ::: : :: |  :|  | :|:| | : |: : || |
Db        857 GSSCCGSGGCGSGCGSGSSGGGSSSSGSGSSGGCGSGSGGGSCSGGGGSCCSCSCGCSCG 916

Qy       1140 CCGGAACTCACCCGCGGCC 1158
              |::   : : ||| ||:||
Db        917 CSSCSSSCSCCCCCCGSCC 935


RESULT 12
BZ569386
LOCUS       BZ569386                1610 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_8365.x1u1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_8365, genomic survey sequence.
ACCESSION   BZ569386
VERSION     BZ569386.1  GI:27203941
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1610)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1610
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_8365"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."

ORIGIN

  Query Match             6.4%;  Score 79.6;  DB 12;  Length 1610;
  Best Local Similarity   46.0%;  Pred. No. 1.2e-05;
  Matches  324;  Conservative    0;  Mismatches  369;  Indels   11;  Gaps    3;

Qy        510 GTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCCC 569
              | | || ||| ||   | ||   ||| | ||   | |||||    ||   ||  ||  |
Db        250 GGGGGCGGCGCGGCGNGGCGCCCCCGGGGCGCCCGGGCCCCGCGGGGGNCGCGCGCCGCG 309

Qy        570 GGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCT 629
              ||||   | |   | |||| ||  |    ||| |     | | |   |||   |
Db        310 GGCGGCGGCGCGCGCGGCCGCCCCGGCCCCGCCCCCGCCGCGCCGCGCCGCCGCCGG 369

Qy        630 TCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGC 689
              |||| |  |    | |    |  |||    |    |||| |  |    ||||
Db        370 GCCGCGGCGCCCGNGCCCGCGCGGCCGCGCCGGGGGGGCCGCCGCGCCGCCGGCGC 429

Qy        690 GGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGG 749
              |        ||   | | || | |  ||||        ||||  |  |||||   |  |||
Db        430 GCCCGCGGGGCGCGCCGCCCGCCGCGCGG-------CGCGCCCGCCGCCGGCGGGCGCGG 482

Qy        750 CCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGC 809
              ||    || |    || | ||  |  | | | || |||| || ||| ||  | || | |
Db        483 CCCGGGCCGGCGCCGGCNCGCGCCCGGCNCGCGCGCCGCCGGCCCCGGCGNNCGGGGCCC 542

Qy        810 GCTGCT-GATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTT 868
              ||  |  |  | ||||||       |||    | || | ||  || | | | ||
Db        543 GCGCCCCGGGCGCGGCGCNNCGGCGCGCGCGCCNGGGCGCCGCGNGCNCGCCGNGCCCGCGC 602

Qy        869 CGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGA 928
              || ||| |  |     |  | | || ||    ||     || | |    ||
Db        603 GCCCGCCGCGGCCGGCNGCGCCCNNGGGGCCCGCGCCCCCGCGCCCGCCGCCGCGCCCCG 662

Qy        929 TCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTA---CGAGCGGCAGTGGA 985
              |  ||||   |  | |||   |    |||  |  |  | |      || |||||
Db        663 CCGCCCGCCNGGNGCCCCGCGCGCCNCGCGCGGGCCCCGCCGGGCGCGCGCGGCCCCCCG 722

Qy        986 GCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCG 1045
              | || || ||||  | ||  | | |  | |    ||| | ||| |  |  | | | |
Db        723 CCGCCGCGCCCCGCNGCCCCGNGGGCGCGCGCGGGCGCGGGCGGGGCCGGGCCCGGCCGGG 782

Qy       1046 GTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCG 1105
              |  | ||  | |  | ||  || ||  | | | ||| | | | ||| | ||| |  |||
Db        783 GCGNGNGGCGGGGGGGCCGGGGCCGGCGGCGCCCGCCGCGGGCCCGCGGGCGCCGCGCCG 842

Qy       1106 CCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGG 1165
              ||    | ||    ||| |  | |||||    |||      |   |||| ||| ||| |
Db        843 CCGNNGGCCCCGCCGCCGCCCGCCGCGCGGGGGGCCGCCCNGCGGGCGCGCCCGCCGCCG 902

Qy       1166 CCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGCGCACACC 1209
              |||| | | ||  || | || |||    |||  |    | | ||
Db        903 CCGGGGNGGCCNGGCGCGCGCGCGGCAGCGGCGCGCGNCNCGCC 946


RESULT 13
AG063827/c
LOCUS       AG063827                989 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-052L03.F, genomic survey sequence.
ACCESSION   AG063827
VERSION     AG063827.1  GI:16615629
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Pujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 989)
```

```
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .989
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-052L03.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             6.3%;  Score 78.8;  DB 14;  Length 989;
  Best Local Similarity   45.2%;  Pred. No. 1.7e-05;
  Matches  347;  Conservative    0;  Mismatches  414;  Indels    6;  Gaps    3;

Qy        460 CCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCG 519
Db        868 CCGGGGGCGGGGCGCGGCGCCCGCGGCCGGCGCGCGGGCCGGCTGGCGCCGCCGGCGGGG 809

Qy        520 GGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAG 579
Db        808 GGCGTCCGGCGGCTGGCGCCCTGGGGGCGCGCGGCGGCGGGGCGGCCCCCCCCCCGCGCC 749

Qy        580 GAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTG 639
Db        748 TCCGGGCCTCCCGCGCCCCCCCTCGCGCCGGCCCGGCGGGCGCGGGGGCCGCCGCGGCCC 689

Qy        640 CGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGG 699
Db        688 GCGCGCGCNTCGCCCCGCCGCCCGGTCGGCCGGCCCCCCGCCGGCGCTGTGGCGTGGGT 629

Qy        700 AAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAG 759
Db        628 TGCTCGCGGGCGGCCGCCCCCGGGGCGGGCCGGCTCGCCCCGGCTGCGCGGNCCGGCGG 569

Qy        760 ATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATC 819
Db        568 G-GCTGCCGCCNCGGGGCGCGGCGCCCGCGGCCGCCGGGNGGCCCGCGGCTTCCCGGGGC 510

Qy        820 ACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGG---CGACG 876
Db        509 CCTCCCCGGCCGGGCGCCGGCCCCGGCCCCCGCCGCCCGGGCCCCCCCTCCGGCGCGCGC 450

Qy        877 CTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGC 936
Db        449 CGCGCCGCGGCGCGCCCGGGGCCCCCCCCCGCCCGGCCGGGCCCGCCCCCCCCGCCCCCC 390

Qy        937 TACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCC 996
Db        389 CGCCCGCGGCGCGCGCCGCGCGGCCCCCGGGCCCCCGGGCGGGGCCGGGGGGCGGGGGGG 330

Qy        997 GCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCAC 1056
Db        329 CGGCCCGCCGCTTNTGCGCGGGCGCCGGGCGCGCGCGGCCCCGCCCGGNCNCCGGGGGCC 270

Qy       1057 ACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCG 1116
Db        269 NCCCGCCCCCCCCCCCCCCCCGGCGCGCCGCGCGCGCCCCCCGCCCGCCCCCCCCCCCCC 210

Qy       1117 AAGCCCGG--CGCCACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGC 1174
Db        209 CCCCCCGGCCCGCGCCCGCGGCGCCCCCNCCCGCGCCCCGGCGCGCCGCGCCCCGCCCC 150

Qy       1175 CCCTGCTCCCGGGCGTACCCGGGCCCGCGCACACCCCCGCAGCGACG 1221
Db        149 CCCCCCCGCNNGCGGGGNNNGGNCNNCCCNNCNNCNCNNNCGGNNCG 103


RESULT 14
BZ569259
LOCUS       BZ569259               1015 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_8319.x2r5p12 pacs2-164 Pseudomonas aeruginosa genomic
            clone pacs2-164_8319, genomic survey sequence.
ACCESSION   BZ569259
VERSION     BZ569259.1  GI:27203589
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1015)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1015
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_8319"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             6.3%;  Score 78.6;  DB 12;  Length 1015;
  Best Local Similarity   46.4%;  Pred. No. 1.8e-05;
  Matches  345;  Conservative    0;  Mismatches  390;  Indels    9;  Gaps    3;

Qy        497 GGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGG 556
Db        115 GGCGGCCGCGCGCCCGGGGCGGCGCGCCGGGCGGCGCGGCGCCGCGGCNGGGGGGGGGGG 174

Qy        557 CGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGT 616
Db        175 GGGGGGGGGGCGCCGCGCGGGGGGGCGGGGCCGGCCGCAGCCACCCNNGTGNTTGNGGGN 234

Qy        617 TCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGC 676
Db        235 GCGGGNCGGGGGCGGGGGCGGCGGGGGCGGGCGGGGGCGGGGGTGGGCGGGGGGGGGCGG 294

Qy        677 TGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCG 736
Db        295 GCGGCCCGCG-GGGGGGCGCGGCGCGCGCGCCGGGCGGGCGCGCGGCGGGGCCTGGCCGC 353

Qy        737 CCGCGAGTTACGGCCTCACCCAG------ATGGTGCCCCCGATCTTCACCCTCGTCGCCT 790
Db        354 GCGCCCGCCGCCGCCCCGCCCCGCGTGGCCCGGCGCGGCCGGCCGGCGCCCGCGCCGCCG 413
```



RESULT 15
AG072551/c
LOCUS AG072551 1370 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-064B08.F, genomic survey sequence.
ACCESSION AG072551
VERSION AG072551.1 GI:16624353
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1370)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1. .1370
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-064B08.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

Query Match 6.3%; Score 78; DB 14; Length 1370;
Best Local Similarity 43.2%; Pred. No. 2.4e-05;
Matches 338; Conservative 0; Mismatches 434; Indels 11; Gaps 3;



Search completed: August 26, 2006, 01:46:21
Job time : 6693 secs

This Page Blank (uspto)

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2006, 21:52:59 ; Search time 297 Seconds
(without alignments)
7843.537 Million cell updates/sec

Title: US-10-680-860A-13
Perfect score: 1245
Sequence: 1 gtgctccagcgcgtcgatct.....aacccaccccgagaggatga 1245

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /EMC_Celerra_SIDS3/ptodata/2/ina/1_COMB.seq:*
2: /EMC_Celerra_SIDS3/ptodata/2/ina/5_COMB.seq:*
3: /EMC_Celerra_SIDS3/ptodata/2/ina/6A_COMB.seq:*
4: /EMC_Celerra_SIDS3/ptodata/2/ina/6B_COMB.seq:*
5: /EMC_Celerra_SIDS3/ptodata/2/ina/7_COMB.seq:*
6: /EMC_Celerra_SIDS3/ptodata/2/ina/H_COMB.seq:*
7: /EMC_Celerra_SIDS3/ptodata/2/ina/PCTUS_COMB.seq:*
8: /EMC_Celerra_SIDS3/ptodata/2/ina/PP_COMB.seq:*
9: /EMC_Celerra_SIDS3/ptodata/2/ina/RE_COMB.seq:*
10: /EMC_Celerra_SIDS3/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

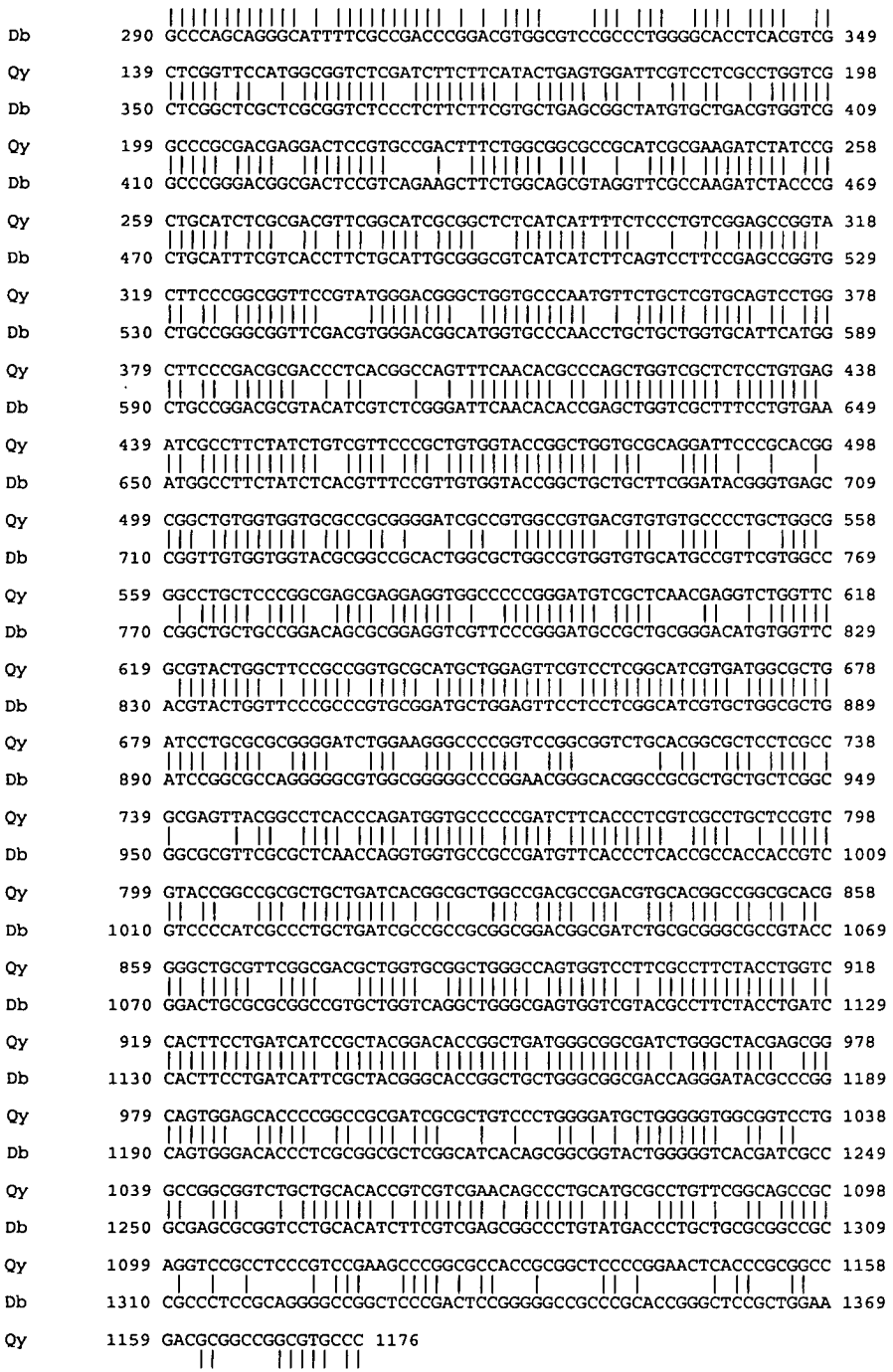
SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	628.4	50.5	1810	2	US-07-708-866A-1	Sequence 1, Appli
	2	626.8	50.3	1810	2	US-07-708-866A-2	Sequence 2, Appli
c	3	355.4	28.5	2749	10	5240849-1	Patent No. 5240849
	4	351.2	28.2	1167	10	5240849-2	Patent No. 5240849
	5	192.8	15.5	47981	3	US-09-679-279-1	Sequence 1, Appli
	6	86.6	7.0	1826	2	US-08-708-856A-14	Sequence 14, Appl
	7	86.6	7.0	1826	3	US-09-287-375-14	Sequence 14, Appl
	8	86.6	7.0	1826	3	US-09-455-406-14	Sequence 14, Appl
	9	86.6	7.0	1826	3	US-10-138-632-14	Sequence 14, Appl
	10	82.4	6.6	3957	3	US-10-237-551-193	Sequence 193, App
	11	82.4	6.6	154746	3	US-09-827-688-8	Sequence 8, Appli
c	12	82.4	6.6	154746	3	US-09-827-688-8	Sequence 8, Appli
	13	66.6	5.3	1317	3	US-09-758-759-46	Sequence 46, Appl
c	14	66.6	5.3	109519	3	US-09-758-759-1	Sequence 1, Appli
	15	64.8	5.2	4257	2	US-08-690-473-1	Sequence 1, Appli
	16	64.8	5.2	4257	3	US-09-259-821A-1	Sequence 1, Appli
	17	64.8	5.2	4257	3	US-08-843-659-1	Sequence 1, Appli
	18	64.8	5.2	4257	3	US-09-825-288A-1	Sequence 1, Appli
c	19	64.8	5.2	12001	2	US-08-458-568A-11	Sequence 11, Appl
	20	64.8	5.2	58857	3	US-09-477-962-1	Sequence 1, Appli
c	21	63.2	5.1	28054	3	US-09-902-540-1248	Sequence 1248, Ap
	22	63.2	5.1	109519	3	US-09-758-759-1	Sequence 1, Appli
	23	63	5.1	5871	3	US-10-152-886-24	Sequence 24, Appl
	24	62	5.0	1839	3	US-09-758-759-150	Sequence 150, App
	25	61	4.9	1263	3	US-09-252-991A-9078	Sequence 9078, Ap
	26	61	4.9	2379	3	US-09-252-991A-9252	Sequence 9252, Ap
	27	61	4.9	11219	2	US-07-642-734C-1	Sequence 1, Appli
	28	61	4.9	11219	3	US-08-439-009A-1	Sequence 1, Appli
c	29	60.2	4.8	942	3	US-09-252-991A-10909	Sequence 10909, A
	30	60.2	4.8	1278	3	US-09-252-991A-11277	Sequence 11277, A
	31	60.2	4.8	1806	3	US-09-252-991A-11213	Sequence 11213, A
	32	59.8	4.8	1215	3	US-09-902-540-5031	Sequence 5031, Ap
	33	59.6	4.8	6225	5	US-09-543-679A-2410	Sequence 2410, Ap
	34	59.6	4.8	35459	5	US-09-543-679A-3003	Sequence 3003, Ap
c	35	59.4	4.8	945	3	US-09-252-991A-10966	Sequence 10966, A
	36	59.4	4.8	30001	2	US-08-125-468-1	Sequence 1, Appli
	37	59.4	4.8	30001	2	US-08-474-933-1	Sequence 1, Appli
c	38	59	4.7	8438	2	US-07-945-283-1	Sequence 1, Appli
	39	58.8	4.7	747	3	US-09-252-991A-695	Sequence 695, App
	40	58.8	4.7	1074	3	US-09-252-991A-722	Sequence 722, App
c	41	58.8	4.7	1266	3	US-09-252-991A-688	Sequence 688, App
	42	58.6	4.7	1074	3	US-09-902-540-8586	Sequence 8586, Ap
c	43	58.6	4.7	1620	3	US-09-252-991A-5364	Sequence 5364, Ap
	44	58.6	4.7	1629	3	US-09-252-991A-5182	Sequence 5182, Ap
	45	58.6	4.7	2067	3	US-09-252-991A-5209	Sequence 5209, Ap

ALIGNMENTS

```
RESULT 1
US-07-708-866A-1
; Sequence 1, Application US/07708866A
; Patent No. 5322937
; GENERAL INFORMATION:
;   APPLICANT: Akira ARISAWA et al.
;   TITLE OF INVENTION: Genes Encoding A 3-Acylation
;   TITLE OF INVENTION: Enzyme For Macrolide Antibodies
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Wenderoth, Lind & Ponack
;     STREET: 805 Fifteenth Street, N.W., #700
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: U.S.A.
;     ZIP: 20005
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: MS-DOS
;     SOFTWARE: DisplayWrite
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/07/708,866A
;     FILING DATE: 19910531
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Warren M. Cheek, Jr.
;     REGISTRATION NUMBER: 33,367
;     REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-371-8850
;     TELEFAX: 202-371-8856
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1810 base pairs
;     TYPE: NUCLEIC ACID
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
;   HYPOTHETICAL: No
;   ANTI-SENSE: No
;   ORIGINAL SOURCE:
;     ORGANISM: Streptomyces thermotolerans
;     STRAIN: ATCC 11416
;   FEATURE:
;     NAME/KEY: -35 signal
```

```
;       LOCATION:  120..125
;       IDENTIFICATION METHOD:  S
;     FEATURE:
;       NAME/KEY:  -10 signal
;       LOCATION:  143..148
;       IDENTIFICATION METHOD:  s
;     FEATURE:
;       NAME/KEY:  RBS
;       LOCATION:  194..198 AND 201..205
;       IDENTIFICATION METHOD:  S
US-07-708-866A-1

  Query Match             50.5%;  Score 628.4;  DB 2;  Length 1810;
  Best Local Similarity   71.4%;  Pred. No. 6.9e-101;
  Matches  827;  Conservative    0;  Mismatches  331;  Indels    0;  Gaps    0;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
Db        230 CTACCTTCACTGACCGGGCTGCGCTGGTTCGCGGCGCTTTCCGTATTCGTCTGCCATATC 289

Qy         79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCCCG 138
Db        290 GCCCAGCAGGGCATTTTCGCCGACCCGGACGTGGCGTCCGCCCTGGGGCACCTCACGTCG 349

Qy        139 CTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCG 198
Db        350 CTCGGCTCGCTCGCGGTCTCCCTCTTCTTCGTGCTGAGCGGCTATGTGCTGACGTGGTCG 409

Qy        199 GCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCG 258
Db        410 GCCCGGGACGGCGACTCCGTCAGAAGCTTCTGGCAGCGTAGGTTCGCCAAGATCTACCCG 469

Qy        259 CTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCGGTA 318
Db        470 CTGCATTTCGTCACCTTCTGCATTGCGGGCGTCATCATCTTCAGTCTTTCCGAGCCGGTG 529

Qy        319 CTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCCTGG 378
Db        530 CTGCCGGGCGGTTCGACGTGGGACGGCATGGTGCCCAACCTGCTGCTGGTGCATTCATGG 589

Qy        379 CTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGTGAG 438
Db        590 CTGCCGGACGCGTACATCGTCTCGGGATTCAACACACCGAGCTGGTCGCTTTCCTGTGAA 649

Qy        439 ATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGG 498
Db        650 ATGGCCTTCTATCTCACGTTTCCGTTGTGGTACCGGCTGCTGCTTCGGATACGGGTGAGC 709

Qy        499 CGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCG 558
Db        710 CGGTTGTGGTGGTACGCGGCCGCACTGGCGCTGGCCGTGGTGTGCATGCCGTTCGTGGCC 769

Qy        559 GGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTC 618
Db        770 CGGCTGCTGCCGGACAGCGCGGAGGTCGTTCCCGGGATGCCGCTGCGGGACATGTGGTTC 829

Qy        619 GCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTG 678
Db        830 ACGTACTGGTTCCCGCCCGTGCGGATGCTGGAGTTCCTCCTCGGCATCGTGCTGGCGCTG 889

Qy        679 ATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCC 738
Db        890 ATCCGGCGCCAGGGGGCGTGGCGGGGGCCCGGAACGGGCACGGCCGCGCTGCTGCTCGGC 949

Qy        739 GCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTC 798
Db        950 GGCGCGTTCGCGCTCAACCAGGTGGTGCCGCCGATGTTCACCCTCACCGCCACCACCGTC 1009

Qy        799 GTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACG 858
Db       1010 GTCCCCATCGCCCTGCTGATCGCCGCCGCGGCGGACGGCGATCTGCGCGGGCGCCGTACC 1069

Qy        859 GGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTC 918
Db       1070 GGACTGCGCGCGGCCGTGCTGGTCAGGCTGGGCGAGTGGTCGTACGCCTTCTACCTGATC 1129

Qy        919 CACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGG 978
Db       1130 CACTTCCTGATCATTCGCTACGGGCACCGGCTGCTGGGCGGCGACCAGGGATACGCCCGG 1189

Qy        979 CAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTG 1038
Db       1190 CAGTGGGACACCCTCGCGGCGCTCGGCATCACAGCGGCGGTACTGGGGGTCACGATCGCC 1249

Qy       1039 GCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGC 1098
Db       1250 GCGAGCGCGGTCCTGCACATCTTCGTCGAGCGGCCCTGTATGACCCTGCTGCGCGGCCGC 1309

Qy       1099 AGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCC 1158
Db       1310 CGCCCTCCGCAGGGGCCGGCTCCCGACTCCGGGGGCCGCCCGCACCGGGCTCCGCTGGAA 1369

Qy       1159 GACGCGGCCGGCGTGCCC 1176
Db       1370 AGGGCATGACGCGTGGCC 1387


RESULT 2
US-07-708-866A-2
; Sequence 2, Application US/07708866A
; Patent No. 5322937
;  GENERAL INFORMATION:
;    APPLICANT:  Akira ARISAWA et al.
;    TITLE OF INVENTION:  Genes Encoding A 3-Acylation
;    TITLE OF INVENTION:  Enzyme For Macrolide Antibodies
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 Kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  DisplayWrite
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/708,866A
;      FILING DATE:  19910531
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:  202-371-8856
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1810 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
US-07-708-866A-2

  Query Match             50.3%;  Score 626.8;  DB 2;  Length 1810;
  Best Local Similarity   71.3%;  Pred. No. 1.3e-100;
  Matches  826;  Conservative    0;  Mismatches  332;  Indels    0;  Gaps    0;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
Db        230 CTACCTTCACTGACCGGGCTGCGCTGGTTCGCGGCGCTTTCCGTATTCGTCTGCCATATC 289

Qy         79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCCCG 138
```



Db 290 GCCCAGCAGGGCATTTTCGCCGACCCGGACGTGGCGTCCGCCCTGGGGCACCTCACGTCG 349

Qy 139 CTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCG 198

Db 350 CTCGGCTCGCTCGCGGTCTCCCTCTTCTTCGTGCTGAGCGGCTATGTGCTGACGTGGTCG 409

Qy 199 GCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCG 258

Db 410 GCCCGGGACGGCGACTCCGTCAGAAGCTTCTGGCAGCGTAGGTTCGCCAAGATCTACCCG 469

Qy 259 CTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCGGTA 318

Db 470 CTGCATTTCGTCACCTTCTGCATTGCGGGCGTCATCATCTTCAGTCCTTCCGAGCCGGTG 529

Qy 319 CTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCCTGG 378

Db 530 CTGCCGGGCGGTTCGACGTGGGACGGCATGGTGCCCAACCTGCTGCTGGTGCATTCATGG 589

Qy 379 CTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGTGAG 438

Db 590 CTGCCGGACGCGTACATCGTCTCGGGATTCAACACACCGAGCTGGTCGCTTTCCTGTGAA 649

Qy 439 ATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGG 498

Db 650 ATGGCCTTCTATCTCACGTTTCCGTTGTGGTACCGGCTGCTGCTTCGGATACGGGTGAGC 709

Qy 499 CGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCG 558

Db 710 CGGTTGTGGTGGTACGCGGCCGCACTGGCGCTGGCCGTGGTGTGCATGCCGTTCGTGGCC 769

Qy 559 GGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTC 618

Db 770 CGGCTGCTGCCGGACAGCGCGGAGGTCGTTCCCGGGATGCCGCTGCGGGACATGTGGTTC 829

Qy 619 GCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTG 678

Db 830 ACGTACTGGTTCCCGCCCGTGCGGATGCTGGAGTTCCTCCTCGGCATCGTGCTGGCGCTG 889

Qy 679 ATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCC 738

Db 890 ATCCGGCGCCAGGGGGCGTGGCGGGGGCCCGGAACGGGCACGGCCGCGCTGCTGCTCGGC 949

Qy 739 GCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTC 798

Db 950 GGCGCGTTCGCGCTCAACCAGGTGGTGCCGCCGATGTTCACCCTCACCGCCACCACCGTC 1009

Qy 799 GTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACG 858

Db 1010 GTCCCCATCGCCCTGCTGATCGCCGCCGCGGCGGACGGCGATCTGCGCGGGCGCCGTACC 1069

Qy 859 GGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTC 918

Db 1070 GGACTGCGCGCGGCCGTGCTGGTCAGGCTGGGCGAGTGGTCGTACGCCTTCTACCTGATC 1129

Qy 919 CACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGG 978

Db 1130 CACTTCCTGATCATTCGCTACGGGCACCGGCTGCTGGGCGGCGACCAGGGATACGCCCGG 1189

Qy 979 CAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTG 1038

Db 1190 CAGTGGGACACCCTCGCGGCGCTCGGCATCACAGCGGCGGTACTGGGGGTCACGATCGCC 1249

Qy 1039 GCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGC 1098

Db 1250 GCGAGCGCGGTCCTGCACATCTTCGTCGAGCGGCCCTGTATGACCCTGCTGCGCGGCCGC 1309

Qy 1099 AGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCC 1158

Db 1310 CGCCCTCCGCAGGGGCCGGCTCCCGACTCCGGGGGCCGCCCGCACCGGGCTCCGCTGGAA 1369

Qy 1159 GACGCGGCCGGCGTGCCC 1176

Db 1370 AGGGCATGACGCGTGGCC 1387

RESULT 3
5240849-1/c
;Patent No. 5240849
; APPLICANT: Arisawa, Akira;Kawamura, Naoto;Kojima,Ikou;Okumura,
;Yasushi;Kazuhiko, Okamura;Hiroshi, Tone;Okamura Rokuro
; TITLE OF INVENTION: DNA CODING FOR ENZYME CAPABLE OF ACLYATING
;THE "4"-POSITION OF MACROLIDE ANTIBIOTIC
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/06/353,323
; FILING DATE: 23-MAY-1989
;SEQ ID NO:1:
; LENGTH: 2749
5240849-1

Query Match 28.5%; Score 355.4; DB 10; Length 2749;
Best Local Similarity 58.2%; Pred. No. 2.3e-53;
Matches 703; Conservative 0; Mismatches 491; Indels 13; Gaps 4;

Qy 2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61

Db 1244 TGCCCCTGCCGAAACATCTTCCCGCGCTCGGCGGGATGCGTTTCATCTCCGCTCTACTGG 1185

Qy 62 TATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCAC 121

Db 1184 TATTCACCTCCCATATATCGACACAGCCGTTCTTCAAGAACACCGAGATCAATTCCGCGC 1125

Qy 122 TGC---TGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTG 178

Db 1124 TGCAGTTCCCGCTGAACCGGCTGGGCCCGCTGACGGTCTCGTTCTTCTTCATGCTCAGCG 1065

Qy 179 GATTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCC 238

Db 1064 GTTTCGTCCTCACCTGGGCGGGTCTGCCCGACAAGTCCAAGGTGAACTTCTGGCGGCGGC 1005

Qy 239 GCATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTT 298

Db 1004 GCACGGTCCGCGCGTACTCGCTGCACCTGCCCGTGCTGCTGGTGACGCTGCTGATCGTGC 945

Qy 299 TCTCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATG 358

Db 944 TGGCCCTCAACGAGCCCAACATG---GGCCGATCGGTGTGGGACGGACTGCTCACGAACC 888

Qy 359 TTCTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCA 418

Db 887 TGCTGCTGATCCAGGCATGGTTCCCCGACCACCACGAGTACGGCAGCATGAACCCGGTGG 828

Qy 419 GCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGG 478

Db 827 CGTGGTCGCTCTCCTGCGAGCTGTTCTTCTACGCCATGTTCCCGTTCCTCTTCGCCTTCT 768

Qy 479 TGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGA 538

Db 767 TCACCAAGGTCCGTACGGACCGGCTCTGGCGGTGGGCCGCCGCGGTGTCCGTGGCCGCCG 708

Qy 539 CGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAG------CGAGGAGGTGGCCCCCG 592

Db 707 TCTCCATCCCCCTGGTCGCACTGCTGCTGCCGGCCAGCCCGCCCCTGCCGTGGGACCCGG 648

Qy 593 GGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGT 652

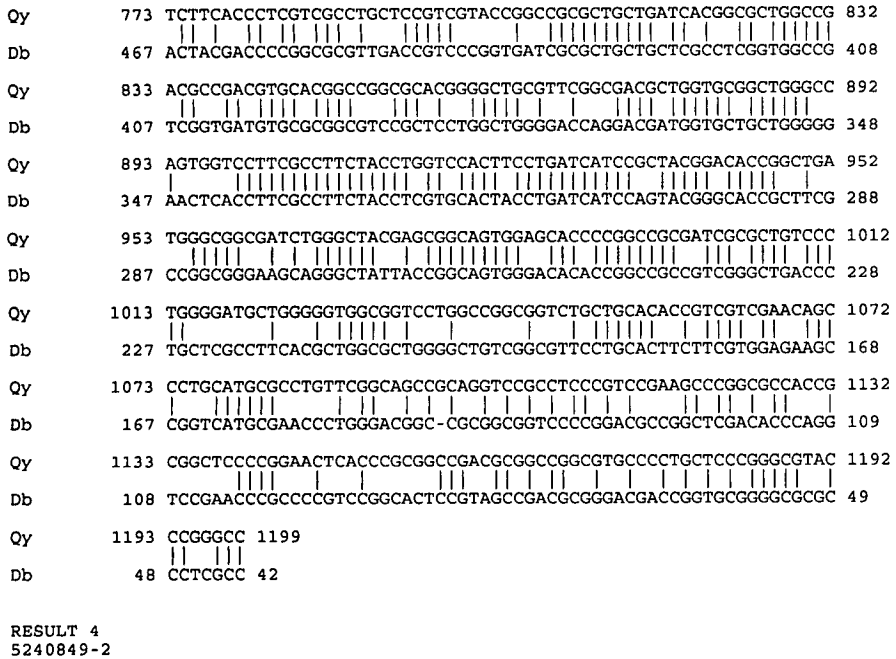
Db 647 ACATGCCGCAGCTGCGGTGGTGGTTCATCTACATGTTCCCGCCGGTGCGGCTGCTGGAGT 588

Qy 653 TCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712

Db 587 TCGTGCTCGGGATGCTCATGGCCCAGATCGTGATCCGGGGACGCTGGAGGGGCCCGCGTC 528

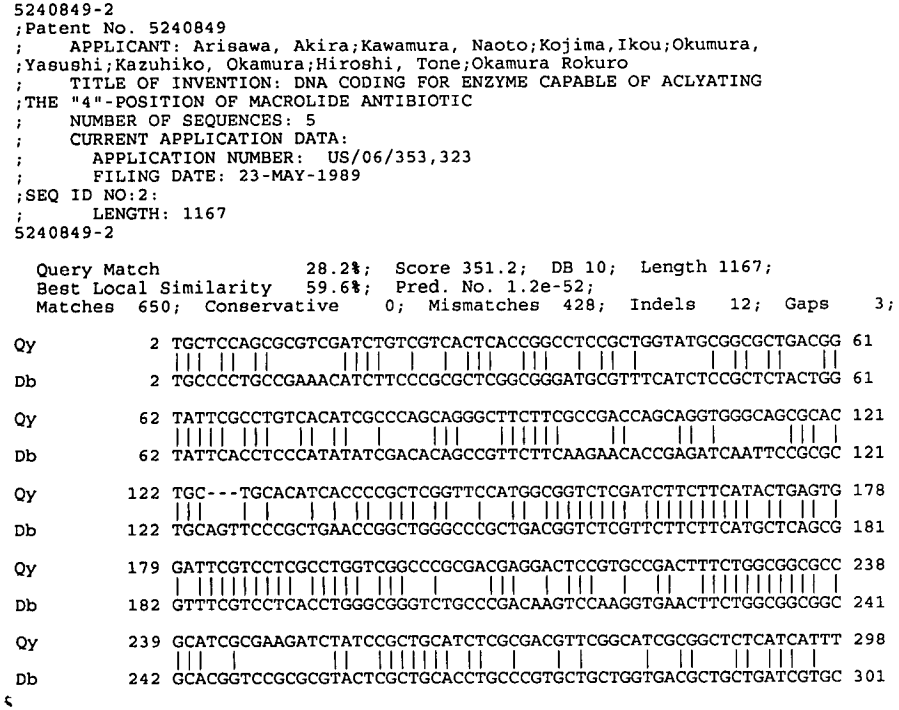
Qy 713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772

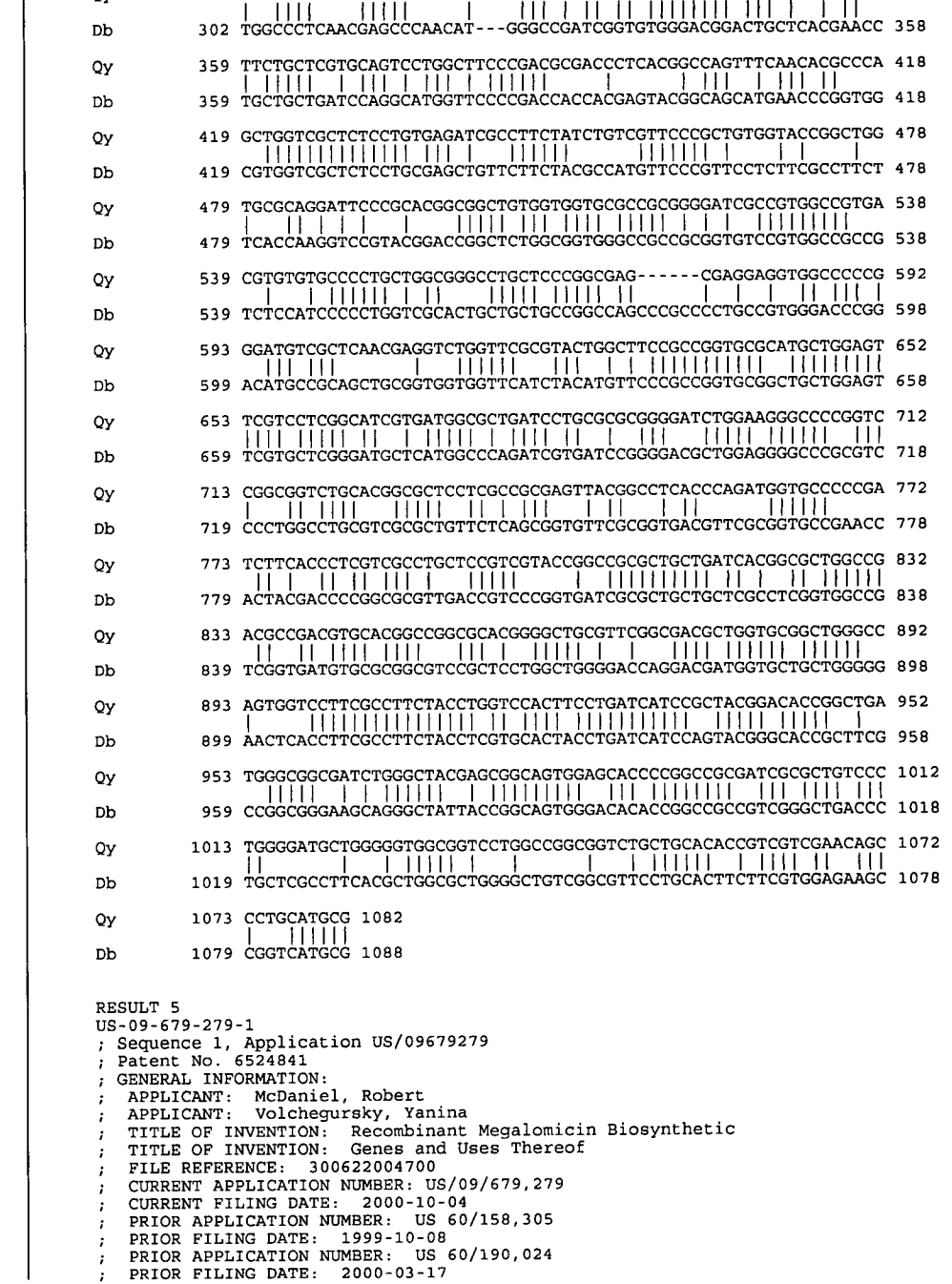
Db 527 CCCTGGCCTGCGTCGCGCTGTTCTCAGCGGTGTTCGCGGTGACGTTCGCGGTGCCGAACC 468



RESULT 4
5240849-2
;Patent No. 5240849
; APPLICANT: Arisawa, Akira;Kawamura, Naoto;Kojima,Ikou;Okumura,
;Yasushi;Kazuhiko, Okamura;Hiroshi, Tone;Okamura Rokuro
; TITLE OF INVENTION: DNA CODING FOR ENZYME CAPABLE OF ACLYATING
;THE "4"-POSITION OF MACROLIDE ANTIBIOTIC
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/06/353,323
; FILING DATE: 23-MAY-1989
;SEQ ID NO:2:
; LENGTH: 1167
5240849-2

Query Match 28.2%; Score 351.2; DB 10; Length 1167;
Best Local Similarity 59.6%; Pred. No. 1.2e-52;
Matches 650; Conservative 0; Mismatches 428; Indels 12; Gaps 3;





RESULT 5
US-09-679-279-1
; Sequence 1, Application US/09679279
; Patent No. 6524841
; GENERAL INFORMATION:
; APPLICANT: McDaniel, Robert
; APPLICANT: Volchegursky, Yanina
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
; TITLE OF INVENTION: Genes and Uses Thereof
; FILE REFERENCE: 300622004700
; CURRENT APPLICATION NUMBER: US/09/679,279
; CURRENT FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/158,305
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 60/190,024
; PRIOR FILING DATE: 2000-03-17

; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 47981
; TYPE: DNA
; ORGANISM: Micromonospora megalomicea
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(144)
; OTHER INFORMATION: megBVI(megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
; OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (928)...(2061)
; OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (2072)...(3382)
; OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),
; OTHER INFORMATION: TDP-megosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (3462)...(4634)
; OTHER INFORMATION: megG(megY), mycarosyl acyltransferase, mycarose O-acyltransferase
; OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (4651)...(5775)
; OTHER INFORMATION: megDII, deoxysugar transaminase (eryCI, DnrJ homolog),
; OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;
; OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (5822)...(6595)
; OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
; OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (6592)...(7197)
; OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmU ho
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
; OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (7220)...(8206)
; OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dnmV homolog),
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
; OTHER INFORMATION: SEQ ID NO NO: 9= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (8228)...(9220)
; OTHER INFORMATION: megBII-1(megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
; OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (9226)...(10479)
; OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (10483)...(11424)
; OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
; OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (12181)...(22821)
; OTHER INFORMATION: megAI; SEQ ID NO: 13= translated amino acid sequence
; NAME/KEY: misc_feature
; LOCATION: (12505)...(13470)
; OTHER INFORMATION: megAI, AT-L
; NAME/KEY: misc_feature
; LOCATION: (13576)...(13791)
; OTHER INFORMATION: megAI, ACP-L
; NAME/KEY: misc_feature
; LOCATION: (13849)...(15126)
; OTHER INFORMATION: megAI, KS1
; NAME/KEY: misc_feature
; LOCATION: (15427)...(16476)
; OTHER INFORMATION: megAI, AT1
; NAME/KEY: misc_feature
; LOCATION: (17155)...(17694)
; OTHER INFORMATION: megAI, KR1
; NAME/KEY: misc_feature
; LOCATION: (17947)...(18207)
; OTHER INFORMATION: megAI, ACP1
; NAME/KEY: misc_feature
; LOCATION: (18268)...(19548)
; OTHER INFORMATION: megAI, KS2
; NAME/KEY: misc_feature
; LOCATION: (19876)...(20910)
; OTHER INFORMATION: megAI, AT2
; NAME/KEY: misc_feature
; LOCATION: (21517)...(22053)
; OTHER INFORMATION: megAI, KR2
; NAME/KEY: misc_feature
; LOCATION: (22318)...(22575)
; OTHER INFORMATION: megAI, ACP2
; NAME/KEY: CDS
; LOCATION: (22867)...(33555)
; OTHER INFORMATION: megAII; SEQ ID NO: 14= translated amino acid sequence
; NAME/KEY: misc_feature
; LOCATION: (22957)...(24237)
; OTHER INFORMATION: megAII, KS3
; NAME/KEY: misc_feature
; LOCATION: (24544)...(25581)
; OTHER INFORMATION: megAII, AT3
; NAME/KEY: misc_feature
; LOCATION: (26230)...(26733)
; OTHER INFORMATION: megAII, KR3 (inactive)
; NAME/KEY: misc_feature
; LOCATION: (26998)...(27258)
; OTHER INFORMATION: megAII, ACP3
; NAME/KEY: misc_feature
; LOCATION: (27393)...(28590)
; OTHER INFORMATION: megAII, KS4
; NAME/KEY: misc_feature
; LOCATION: (28897)...(29931)
; OTHER INFORMATION: megAII, AT4
; NAME/KEY: misc_feature
; LOCATION: (29953)...(30477)
; OTHER INFORMATION: megAII, DH4
; NAME/KEY: misc_feature
; LOCATION: (31396)...(32244)
; OTHER INFORMATION: megAII, ER4
; NAME/KEY: misc_feature
; LOCATION: (32257)...(32799)
; OTHER INFORMATION: megAII, KR4
; NAME/KEY: misc_feature
; LOCATION: (33052)...(33312)
; OTHER INFORMATION: megAII, ACP4
; NAME/KEY: CDS
; LOCATION: (33666)...(43271)
; OTHER INFORMATION: megAIII; SEQ ID NO: 15= translated amino acid sequence
; NAME/KEY: misc_feature
; LOCATION: (33780)...(35027)
; OTHER INFORMATION: megAIII, KS5
; NAME/KEY: misc_feature
; LOCATION: (35385)...(36419)
; OTHER INFORMATION: megAIII, AT5
; NAME/KEY: misc_feature
; LOCATION: (37068)...(37604)
; OTHER INFORMATION: megAIII, KR5
; NAME/KEY: misc_feature
; LOCATION: (37860)...(38120)
; OTHER INFORMATION: megAIII, ACP5
; NAME/KEY: misc_feature
; LOCATION: (38187)...(39470)
; OTHER INFORMATION: megAIII, KS6
; NAME/KEY: misc_feature
; LOCATION: (39795)...(40811)
; OTHER INFORMATION: megAIII, AT6
; NAME/KEY: misc_feature
; LOCATION: (41406)...(41936)

```
;   OTHER INFORMATION: megAIII, KR6
;   NAME/KEY: misc_feature
;   LOCATION: (42168)...(42425)
;   OTHER INFORMATION: megAIII, ACP6
;   NAME/KEY: misc_feature
;   LOCATION: (42585)...(43271)
;   OTHER INFORMATION: megAIII, TE
;   NAME/KEY: CDS
;   LOCATION: (43268)...(44344)
;   OTHER INFORMATION: megCII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
;   OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
;   NAME/KEY: CDS
;   LOCATION: (44355)...(45623)
;   OTHER INFORMATION: megCIII, desosaminyl transferase, desosamine glycosyltransferase;
;   OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
;   NAME/KEY: CDS
;   LOCATION: (45620)...(46591)
;   OTHER INFORMATION: megBII-2(megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
;   OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
;   OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
;   NAME/KEY: CDS
;   LOCATION: (46660)...(47403)
;   OTHER INFORMATION: megH, TEII; SEQ ID NO: 19= translated amino acid sequence
;   NAME/KEY: CDS
;   LOCATION: (47411)...(47980)
;   OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid seque
US-09-679-279-1

  Query Match           15.5%;  Score 192.8;  DB 3;  Length 47981;
  Best Local Similarity  53.9%;  Pred. No. 4.8e-25;
  Matches  526;  Conservative    0;  Mismatches  417;  Indels   33;  Gaps    5;

Qy       149 TGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCGGCCCGCGACG 208
Db      3637 TGGGGGTGTCGTTCTTCTTTATTCTCAGCGGTTTCGTGCTGACCTGGTCGGCGCGGGCCA 3696

Qy       209 AGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCGCTGCATCTCG 268
Db      3697 GCGACTCGGTGTGGTCGTTCTGGCGCAGACGGGTCTGCAAGCTCTTCCCCAACCACCTGG 3756

Qy       269 CGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCGGTACTTCCCGGCG 328
Db      3757 TCAC---CGCCTTCGCCGCCGTGGTGTTGTTCCTGGTCACCGGGCAGGCGGTGAGCG--- 3810

Qy       329 GTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCCTGGCTTCCCGACG 388
Db      3811 ---------GTGAGGCGCTGATCCCGAACCTCCTGCTGATCCACGCCTGGTTCCCGGCCC 3861

Qy       389 CGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGTGAGATCGCCTTCT 448
Db      3862 TGGAGATCTCCTTCGGCATCAACCCGGTGAGCTGGTCGTTGGCCTGCGAGGCGTTCTTCT 3921

Qy       449 ATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGGCGGCTGTGGT 508
Db      3922 ACCTGTGCTTCCCGCTGTTCCTGTTCTGGATCTCCGGTATCCGCCCGGAGCGGCTGTGGG 3981

Qy       509 GGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCC 568
Db      3982 CCTGGGCCGCCGTGGTGTTCGCCGCGATCTGGGCGGTACCGGTGGTCGCCGACCTCCTGC 4041

Qy       569 CGGCGAG---------CGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTCG 619
Db      4042 TGCCGAGTTCCCCGCCGCTGATCCCGGGGCTTGAGTACTCCGCCATCCAGGACTGGTTCC 4101

Qy       620 CGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTGA 679
Db      4102 TCTACACCTTCCCTGCGACGCGGAGCCTGGAGTTCATCCTCGGGATCATCCTGGCCCGCA 4161

Qy       680 TCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCG 739
Db      4162 TCCTGATCACCGGTCGGTGGATCAACGTCGGGCTGCTCCCCGCGGTGCTGTTGTTCCCGG 4221

Qy       740 CGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCG 799
Db      4222 TCTTCTTCGTCGCCTCGCTCTTCCTGCCGGGTGTCTACGCCATCTCCTCGTCGATGATGA 4281

Qy       800 TACCGGCCGCGCTGCTGATCACGGCGCTGGCCGAC---GCCGACGTGCACGGCCGGCGCA 856
Db      4282 TCCTTCCCCTGGTTCTGATCATCGCCAGCGGCGCGACGGCCGACCTCCAGCAGAAGCGCA 4341

Qy       857 CGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGG 916
Db      4342 CCTTCATGCGTAACCGGGTGATGGTGTGGCTCGGCGACGTCTCCTTCGCGCTCTACATGG 4401

Qy       917 TCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGC 976
Db      4402 TCCACTTCCTGGTGATCGTCTACGGGGCGGACCTGCTGGGGTTCAGCCAGACCGAGGACG 4461

Qy       977 GGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCC 1036
Db      4462 CCCCGCTGGG------TCTCGCACTCTTCATGATCATTCCGTTCCTCGCGGTCTCCCTGG 4515

Qy      1037 TGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCC 1096
Db      4516 TGCTGTCGTGGCTGCTGTACAGGTTCGTCGAGCTACCCGTCATGCGTAACTGGGCCCGCC 4575

Qy      1097 GCAGGTCCGCCTCCCG 1112
Db      4576 CGGCCTCCGCCCGGCG 4591


RESULT 6
US-08-708-856A-14
; Sequence 14, Application US/08708856A
; Patent No. 5963254
;  GENERAL INFORMATION:
;    APPLICANT:  Ozaki, Akio
;    APPLICANT:  Mori, Hideo
;    APPLICANT:  Shibasaki, Takeshi
;    APPLICANT:  Ando, Katsuhiko
;    APPLICANT:  Ochiai, Keiko
;    APPLICANT:  Chiba, Shigeru
;    APPLICANT:  Uosaki, Youichi
;    TITLE OF INVENTION:  Process for Producing
;    TITLE OF INVENTION:  Cis-3-Hydroxy-L-Proline
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ANTONELLI, TERRY, STOUT AND KRAUS
;      STREET:  1300 NORTH SEVENTEENTH STREET
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22209
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/708,856A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/301,654
;      FILING DATE:  07-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/474,135
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terry, David T.
;      REGISTRATION NUMBER:  20178
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-312-6600
;      TELEFAX:  703-312-6666
```

```
;      TELEX:  248545
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1826 base paires
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces sp.
;      STRAIN:  TH1
US-08-708-856A-14

  Query Match             7.0%;  Score 86.6;  DB 2;  Length 1826;
  Best Local Similarity   51.9%;  Pred. No. 1.4e-06;
  Matches  296;  Conservative    0;  Mismatches  235;  Indels   39;  Gaps    3;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
Db       1245 CTGCCCTCGCTGACCGGCCTGCGCTTCCCGGCGGCGTTCATCGTGTTCCTCTTCCACGCC 1304

Qy         79 GCCCAGCAG---------------GGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTG 123
Db       1305 TCGCTGCCGTTTCCCGCGGTGCGCCTGTTCGCCGACGACGGGGTGGAACACCGCTACGGG 1364

Qy        124 CTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTC 183
Db       1365 TGGGCCCTCGGCCCGAGCGGCGCACTCGGCGTGACGTTCTTCTTCGCGCTCAGCGGATTC 1424

Qy        184 GTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATC 243
Db       1425 GTGCTGACGTGGTCGGCTCCCGCCGGCGACACCGCACCGTCCTTCTGGCGCCGGCGCTTC 1484

Qy        244 GCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCC 303
Db       1485 GTCAAGATCCTGCCCAACTACGTCGTCGCGTGGGTCCTGGC------------------- 1525

Qy        304 CTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTG 363
Db       1526 --GATGGTGCTGTACGCGGCCGCGACGCCCGTCCTGCCCGCGCTCGGCGCCCTTTTCATG 1583

Qy        364 CTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGG 423
Db       1584 CTC---CAGGTGTGGACGCCGTACTTCACCGAGCACCTCCCGGTGAACCCACCGAGCTGG 1640

Qy        424 TCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGC 483
Db       1641 TCGCTGNCCGTGGAAGCCGTCTTCTATCTGGCCTTCCCGTTCCTGCTGGCCGGGATCAGA 1700

Qy        484 AGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGT 543
Db       1701 CGGATACCGGCCGCCCGGCTGAAGTACTGGATCGCCGGCACGGTGGCCGCCGTCTTCGCC 1760

Qy        544 GTGCCCCTGCTGGCGGGCCTGCTCCCGGCG 573
Db       1761 ACGCCGCTGATCACCTACCTCCTGGTGCCG 1790


RESULT 7
US-09-287-375-14
; Sequence 14, Application US/09287375
; Patent No. 6010891
;  GENERAL INFORMATION:
;    APPLICANT:  Ozaki, Akio
;    APPLICANT:  Mori, Hideo
;    APPLICANT:  Shibasaki, Takeshi
;    APPLICANT:  Ando, Katsuhiko
;    APPLICANT:  Ochiai, Keiko
;    APPLICANT:  Chiba, Shigeru
;    APPLICANT:  Uosaki, Youichi
;    TITLE OF INVENTION:  Process for Producing
;    TITLE OF INVENTION:  Cis-3-Hydroxy-L-Proline
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ANTONELLI, TERRY, STOUT AND KRAUS
;      STREET:  1300 NORTH SEVENTEENTH STREET
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22209
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/287,375
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/708,856
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/474,135
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terry, David T.
;      REGISTRATION NUMBER:  20178
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-312-6600
;      TELEFAX:  703-312-6666
;      TELEX:  248545
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1826 base paires
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces sp.
;      STRAIN:  TH1
US-09-287-375-14

  Query Match             7.0%;  Score 86.6;  DB 3;  Length 1826;
  Best Local Similarity   51.9%;  Pred. No. 1.4e-06;
  Matches  296;  Conservative    0;  Mismatches  235;  Indels   39;  Gaps    3;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
Db       1245 CTGCCCTCGCTGACCGGCCTGCGCTTCCCGGCGGCGTTCATCGTGTTCCTCTTCCACGCC 1304

Qy         79 GCCCAGCAG---------------GGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTG 123
Db       1305 TCGCTGCCGTTTCCCGCGGTGCGCCTGTTCGCCGACGACGGGGTGGAACACCGCTACGGG 1364

Qy        124 CTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTC 183
Db       1365 TGGGCCCTCGGCCCGAGCGGCGCACTCGGCGTGACGTTCTTCTTCGCGCTCAGCGGATTC 1424

Qy        184 GTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATC 243
Db       1425 GTGCTGACGTGGTCGGCTCCCGCCGGCGACACCGCACCGTCCTTCTGGCGCCGGCGCTTC 1484

Qy        244 GCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCC 303
Db       1485 GTCAAGATCCTGCCCAACTACGTCGTCGCGTGGGTCCTGGC------------------- 1525

Qy        304 CTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTG 363
Db       1526 --GATGGTGCTGTACGCGGCCGCGACGCCCGTCCTGCCCGCGCTCGGCGCCCTTTTCATG 1583

Qy        364 CTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGG 423
```

```
Db       1584 CTC---CAGGTGTGGACGCCGTACTTCACCGAGCACCTCCCGGTGAACCCACCGAGCTGG 1640

Qy        424 TCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGC 483
              |||||  ||    ||   || |||||||||| | ||||||  |   |         | |  |
Db       1641 TCGCTGNCCGTGGAAGCCGTCTTCTATCTGGCCTTCCCGTTCCTGCTGGCCGGGATCAGA 1700

Qy        484 AGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGT 543
              ||||  || ||    ||||||  ||  ||   ||| || |  |  |  |||||
Db       1701 CGGATACCGGCCGCCCGGCTGAAGTACTGGATCGCCGGCACGGTGGCCGCCGTCTTCGCC 1760

Qy        544 GTGCCCCTGCTGGCGGGCCTGCTCCCGGCG 573
              ||| ||| |  |   ||| ||   | ||
Db       1761 ACGCCGCTGATCACCTACCTCCTGGTGCCG 1790


RESULT 8
US-09-455-406-14
; Sequence 14, Application US/09455406
; Patent No. 6413748
;  GENERAL INFORMATION:
;    APPLICANT:  Ozaki, Akio
;    APPLICANT:  Mori, Hideo
;    APPLICANT:  Shibasaki, Takeshi
;    APPLICANT:  Ando, Katsuhiko
;    APPLICANT:  Ochiai, Keiko
;    APPLICANT:  Chiba, Shigeru
;    APPLICANT:  Uosaki, Youichi
;    TITLE OF INVENTION:  Process for Producing
;    TITLE OF INVENTION:  Cis-3-Hydroxy-L-Proline
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ANTONELLI, TERRY, STOUT AND KRAUS
;      STREET:  1300 NORTH SEVENTEENTH STREET
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22209
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/455,406
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/708,856
;      FILING DATE:
;      APPLICATION NUMBER:  08/301,654
;      FILING DATE:  07-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/474,135
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terry, David T.
;      REGISTRATION NUMBER:  20178
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-312-6600
;      TELEFAX:  703-312-6666
;      TELEX:  248545
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1826 base paires
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces sp.
;      STRAIN:  TH1
US-09-455-406-14
```

```
  Query Match             7.0%;  Score 86.6;  DB 3;  Length 1826;
  Best Local Similarity   51.9%;  Pred. No. 1.4e-06;
  Matches  296;  Conservative    0;  Mismatches  235;  Indels   39;  Gaps    3;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
              ||| | || || |||||||| ||||      ||||| | |  || |||  ||  |||  |
Db       1245 CTGCCCTCGCTGACCGGCCTGCGCTTCCCGGCGGCGTTCATCGTGTTCCTCTTCCACGCC 1304

Qy         79 GCCCAGCAG----------------GGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTG 123
              | | || |                    |||  || ||||  |  ||    | ||      |
Db       1305 TCGCTGCCGTTTCCCGCGGTGCGCCTGTTCGCCGACGACGGGGTGGAACACCGCTACGGG 1364

Qy        124 CTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTC 183
              |  | ||  ||||  |||  |  | |  ||  || |||||||||    || || ||||||
Db       1365 TGGGCCCTCGGCCCGAGCGGCGCACTCGGCGTGACGTTCTTCTTCGCGCTCAGCGGATTC 1424

Qy        184 GTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATC 243
              || ||  | |||||||| | || ||  ||| |||  ||| | ||||||||| || ||| ||
Db       1425 GTGCTGACGTGGTCGGCTCCCGCCGGCGACACCGCACCGTCCTTCTGGCGCCGGCGCTTC 1484

Qy        244 GCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCC 303
              |  ||||||   ||      | |||   |||  | | | ||
Db       1485 GTCAAGATCCTGCCCAACTACGTCGTCGCGTGGGTCCTGGC------------------- 1525

Qy        304 CTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTG 363
              |  || || |       |||       ||||   |  || ||| |   ||  | |  ||
Db       1526 --GATGGTGCTGTACGCGGCCGCGACGCCCGTCCTGCCCGCGCTCGGCGCCCTTTTCATG 1583

Qy        364 CTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGG 423
              |||  |||  |||  ||  ||  |||      |   | ||| | || ||||||
Db       1584 CTC---CAGGTGTGGACGCCGTACTTCACCGAGCACCTCCCGGTGAACCCACCGAGCTGG 1640

Qy        424 TCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGC 483
              |||||  ||    ||   || |||||||||| | ||||||  |   |         | |  |
Db       1641 TCGCTGNCCGTGGAAGCCGTCTTCTATCTGGCCTTCCCGTTCCTGCTGGCCGGGATCAGA 1700

Qy        484 AGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGT 543
              ||||  || ||    ||||||  ||  ||   ||| || |  |  |  |||||
Db       1701 CGGATACCGGCCGCCCGGCTGAAGTACTGGATCGCCGGCACGGTGGCCGCCGTCTTCGCC 1760

Qy        544 GTGCCCCTGCTGGCGGGCCTGCTCCCGGCG 573
              ||| ||| |  |   ||| ||   | ||
Db       1761 ACGCCGCTGATCACCTACCTCCTGGTGCCG 1790


RESULT 9
US-10-138-632-14
; Sequence 14, Application US/10138632
; Patent No. 6767726
;   GENERAL INFORMATION:
;        APPLICANT: Ozaki, Akio
;                   Mori, Hideo
;                   Shibasaki, Takeshi
;                   Ando, Katsuhiko
;                   Ochiai, Keiko
;                   Chiba, Shigeru
;                   Uosaki, Youichi
;        TITLE OF INVENTION: Process for Producing
;                            Cis-3-Hydroxy-L-Proline
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: ANTONELLI, TERRY, STOUT AND KRAUS
;             STREET: 1300 NORTH SEVENTEENTH STREET
;             CITY: ARLINGTON
;             STATE: VIRGINIA
;             COUNTRY: U.S.A.
;             ZIP: 22209
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
```

```
;             APPLICATION NUMBER: US/10/138,632
;             FILING DATE: 06-May-2002
;             CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/455,406
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/708,856
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/301,654
;             FILING DATE: 07-SEP-1994
;             APPLICATION NUMBER: 08/474,135
;             FILING DATE: 07-JUN-1995
;       ATTORNEY/AGENT INFORMATION:
;             NAME: Terry, David T.
;             REGISTRATION NUMBER: 20178
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 703-312-6600
;             TELEFAX: 703-312-6666
;             TELEX: 248545
;   INFORMATION FOR SEQ ID NO: 14:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 1826 base paires
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;       MOLECULE TYPE: genomic DNA
;       ORIGINAL SOURCE:
;             ORGANISM: Streptomyces sp.
;             STRAIN: TH1
;       SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-138-632-14

  Query Match             7.0%;  Score 86.6;  DB 3;  Length 1826;
  Best Local Similarity   51.9%;  Pred. No. 1.4e-06;
  Matches  296;  Conservative    0;  Mismatches  235;  Indels   39;  Gaps    3;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
              ||| | || || |||||||| ||||     ||||||  | |  || |||  ||  ||| |
Db       1245 CTGCCCTCGCTGACCGGCCTGCGCTTCCCGGCGGCGTTCATCGTGTTCCTCTTCCACGCC 1304

Qy         79 GCCCAGCAG----------------GGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTG 123
                | | || |                    |||  || ||||  |  ||    | ||     |
Db       1305 TCGCTGCCGTTTCCCGCGGTGCGCCTGTTCGCCGACGACGGGGTGGAACACCGCTACGGG 1364

Qy        124 CTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTC 183
                |  | ||  ||||  |||  |  | |  ||  || ||||||||   || || ||||||
Db       1365 TGGGCCCTCGGCCCGAGCGGCGCACTCGGCGTGACGTTCTTCTTCGCGCTCAGCGGATTC 1424

Qy        184 GTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATC 243
              || ||  | |||||||| | || ||  ||| |||  ||| | |||||||| || ||| ||
Db       1425 GTGCTGACGTGGTCGGCTCCCGCCGGCGACACCGCACCGTCCTTCTGGCGCCGGCGCTTC 1484

Qy        244 GCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCC 303
                |  ||||||   ||     |  |||   |||  | | | ||
Db       1485 GTCAAGATCCTGCCCAACTACGTCGTCGCGTGGGTCCTGGC------------------- 1525

Qy        304 CTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTG 363
                |  || || |       |||      ||||   |  || ||| |   ||  | |  ||
Db       1526 --GATGGTGCTGTACGCGGCCGCGACGCCCGTCCTGCCCGCGCTCGGCGCCCTTTTCATG 1583

Qy        364 CTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGG 423
              |||   |||   |||  ||  ||   |||         |    | ||| | || ||||||
Db       1584 CTC---CAGGTGTGGACGCCGTACTTCACCGAGCACCTCCCGGTGAACCCACCGAGCTGG 1640

Qy        424 TCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGC 483
              |||||  ||   ||   || |||||||||| | |||||| |   |      | | |
Db       1641 TCGCTGNCCGTGGAAGCCGTCTTCTATCTGGCCTTCCCGTTCCTGCTGGCCGGGATCAGA 1700

Qy        484 AGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGT 543
              ||||  || ||    ||||||  ||  ||   ||| || |  |  |  |||||
Db       1701 CGGATACCGGCCGCCCGGCTGAAGTACTGGATCGCCGGCACGGTGGCCGCCGTCTTCGCC 1760

Qy        544 GTGCCCCTGCTGGCGGGCCTGCTCCCGGCG 573
```

```
                ||| ||| |  |   ||| ||   | ||
Db       1761 ACGCCGCTGATCACCTACCTCCTGGTGCCG 1790


RESULT 10
US-10-237-551-193
; Sequence 193, Application US/10237551
; Patent No. 6821519
; GENERAL INFORMATION:
;  APPLICANT: Day, Craig H.
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Parsons, Joseph M.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
;  FILE REFERENCE: 210121.538C3
;  CURRENT APPLICATION NUMBER: US/10/237,551
;  CURRENT FILING DATE:  2002-09-06
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 193
;   LENGTH: 3957
;   TYPE: DNA
;   ORGANISM: HSV2
US-10-237-551-193

  Query Match             6.6%;  Score 82.4;  DB 3;  Length 3957;
  Best Local Similarity   44.6%;  Pred. No. 7.6e-06;
  Matches  458;  Conservative    0;  Mismatches  561;  Indels    9;  Gaps    3;

Qy        183 CGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCAT 242
              || |  |||     | |||  || |  |     ||   ||         ||| |||
Db       2769 CGGCGGCGCGCCCGCCGCCTTCGGCCCGCTGCGCGCCTCGGGCCCGCTGCGCCGCGCGGC 2828

Qy        243 CGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTC 302
                ||   |||    | | |||    | ||  |  ||    || ||   || ||    |   |
Db       2829 GGCCTGGATGCGCCAGGTGCCCGACCCGGAGGACGTGCGCGTGGTGATCCTCTACTCGCC 2888

Qy        303 CCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCT 362
                ||| |||    ||  ||  |||  ||   ||   ||| |||||      ||  |
Db       2889 GCTGCCGGGCGAGGACCTGGCCGCGGGCCGCGCCGGGGGCGGGCCCCCCCCGGAGTGGTC 2948

Qy        363 GCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTG 422
                || ||   | ||||  ||  |  |   ||  |||||     ||||  || | || |
Db       2949 CGCCGAGCGCGGCGGGCTGTCCTGCCTG---CTGGCGGCCCTGGGCAACCGGCTCTGCGG 3005

Qy        423 GTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCG 482
              | |   | |    |    ||   | |   | |   | | |        |  |  | ||| |
Db       3006 GCCCGCCACGGCCGCCTGGGCGGGCAACTGGACCGGCGCCCCCGACGTCTCGGCGCTGGG 3065

Qy        483 CAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTG 542
              |  |      ||  | ||  |||||    | | | |  ||   ||||||  ||||||  ||
Db       3066 CGCGCAGGGCGTGCTGCTGCTGTCCACGCGGGACCTGGCCTTCGCCGGCGCCGTGGAGTT 3125

Qy        543 TGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCT 602
                ||  |||||||| |||      |        ||       || | |  ||   |  ||
Db       3126 CCTGGGGCTGCTGGCCGGCGCCTGCGACCGCCGCCTCATCGTCGTCAACGCCGT---GCG 3182

Qy        603 CAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGG 662
              |  || || ||||  |||  || |||      |      |||   ||  |  | | |  | |
Db       3183 CGCCGCGGACTGGCCCGCCGACGGGCCCGTGGTCTCGCGGCAGCACGCCTACCTGGCCTG 3242

Qy        663 CATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTG 722
              |  ||| || |    | | | |||| | |  |||| ||   |  |       | | |
Db       3243 CGAGGTGCTGCCCGCCGTGCAGTGCGCCGTGCGCTGGCCGGCGGCGCGGGACCTGCGCCG 3302

Qy        723 CACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCT 782
              ||| | |||   | |||    |      |    |  | |  |  | |   |    ||
Db       3303 CACCGTGCTGGCCTCCGGCCGCGTGTTCGGGCCGGGGGTCTTCGCGCGCGTGGAGGCCGC 3362

Qy        783 CGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGT 842
                |||  ||      |      ||  ||| ||||  |       |  |  | |  ||
Db       3363 GCACGCGCGCCTGTACCCCGACGCGCCGCCGCTGCGCCTCTGCCGCGGGGCCAACGTGCG 3422
```

```
Qy        843 GCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTT 902
Db       3423 GTACCGCGTGCGCACGCGCTTCGGCCCCGACACGCTGGTGCCCATGTCCCCGCGCGAGTA 3482

Qy        903 CGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGA 962
Db       3483 CCGCCGCGCCGTGCTCCCGGCGCTGGACGGCCGGGCCGCCGCCTCGGGCGCGGGCGACGC 3542

Qy        963 TCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCT 1022
Db       3543 CATGGCGCCCGGCGCGCCGGACTTCTGCGAGGACGAG---GCGCACTCGCACCGCGCCTG 3599

Qy       1023 GGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCG 1082
Db       3600 CGCGCGCTGGGGCCTGGGCGCGCCGCTGCGGCCCGTCTACGTGGCGCTGGGGCGCGACGC 3659

Qy       1083 CCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCG 1142
Db       3660 CGTGCGCGGCGGCCCGGCGGAGCTGCGCGGGCCGCGGCGGGAGTTCTGCGCGCGGGCGCT 3719

Qy       1143 GAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGC 1202
Db       3720 GCTCGAGCCCGACGGCGACGCGCCCCCGCTGGTGCTGCGCGACGACGCGGACGCGGGCCC 3779

Qy       1203 GCACACCC 1210
Db       3780 GCCCCCGC 3787


RESULT 11
US-09-827-688-8
; Sequence 8, Application US/09827688
; Patent No. 6821955
; GENERAL INFORMATION:
;  APPLICANT: ORSON, FRANK
;  APPLICANT:  KINSEY, BERMA
;  APPLICANT:  BHOGAL, BALBIR
;  TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION D
;  TITLE OF INVENTION:  AGENTS
;  FILE REFERENCE: P01949US1/10004014
;  CURRENT APPLICATION NUMBER: US/09/827,688
;  CURRENT FILING DATE:  2001-04-06
;  PRIOR APPLICATION NUMBER: 60/195,680
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 154746
;   TYPE: DNA
;   ORGANISM: HERPESVIRUS 2
US-09-827-688-8

  Query Match             6.6%;  Score 82.4;  DB 3;  Length 154746;
  Best Local Similarity   44.6%;  Pred. No. 7.9e-06;
  Matches  458;  Conservative    0;  Mismatches  561;  Indels    9;  Gaps    3;

Qy        183 CGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCAT 242
Db     152478 CGGCGGCGCGCCCGCCGCCTTCGGCCCGCTGCGCGCCTCGGGCCCGCTGCGCCGCGCGGC 152537

Qy        243 CGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTC 302
Db     152538 GGCCTGGATGCGCCAGGTGCCCGACCCGGAGGACGTGCGCGTGGTGATCCTCTACTCGCC 152597

Qy        303 CCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCT 362
Db     152598 GCTGCCGGGCGAGGACCTGGCCGCGGGCCGCGCCGGGGGCGGGCCCCCCCCGGAGTGGTC 152657

Qy        363 GCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTG 422
Db     152658 CGCCGAGCGCGGCGGGCTGTCCTGCCTG---CTGGCGGCCCTGGGCAACCGGCTCTGCGG 152714

Qy        423 GTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCG 482
Db     152715 GCCCGCCACGGCCGCCTGGGCGGGCAACTGGACCGGCGCCCCCGACGTCTCGGCGCTGGG 152774

Qy        483 CAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTG 542
Db     152775 CGCGCAGGGCGTGCTGCTGCTGTCCACGCGGGACCTGGCCTTCGCCGGCGCCGTGGAGTT 152834

Qy        543 TGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCT 602
Db     152835 CCTGGGGCTGCTGGCCGGCGCCTGCGACCGCCGCCTCATCGTCGTCAACGCCGT---GCG 152891

Qy        603 CAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGG 662
Db     152892 CGCCGCGGACTGGCCCGCCGACGGGCCCGTGGTCTCGCGGCAGCACGCCTACCTGGCCTG 152951

Qy        663 CATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTG 722
Db     152952 CGAGGTGCTGCCCGCCGTGCAGTGCGCCGTGCGCTGGCCGGCGGCGCGGGACCTGCGCCG 153011

Qy        723 CACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCT 782
Db     153012 CACCGTGCTGGCCTCCGGCCGCGTGTTCGGGCCGGGGGTCTTCGCGCGCGTGGAGGCCGC 153071

Qy        783 CGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGT 842
Db     153072 GCACGCGCGCCTGTACCCCGACGCGCCGCCGCTGCGCCTCTGCCGCGGGGCCAACGTGCG 153131

Qy        843 GCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTT 902
Db     153132 GTACCGCGTGCGCACGCGCTTCGGCCCCGACACGCTGGTGCCCATGTCCCCGCGCGAGTA 153191

Qy        903 CGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGA 962
Db     153192 CCGCCGCGCCGTGCTCCCGGCGCTGGACGGCCGGGCCGCCGCCTCGGGCGCGGGCGACGC 153251

Qy        963 TCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCT 1022
Db     153252 CATGGCGCCCGGCGCGCCGGACTTCTGCGAGGACGAG---GCGCACTCGCACCGCGCCTG 153308

Qy       1023 GGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCG 1082
Db     153309 CGCGCGCTGGGGCCTGGGCGCGCCGCTGCGGCCCGTCTACGTGGCGCTGGGGCGCGACGC 153368

Qy       1083 CCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCG 1142
Db     153369 CGTGCGCGGCGGCCCGGCGGAGCTGCGCGGGCCGCGGCGGGAGTTCTGCGCGCGGGCGCT 153428

Qy       1143 GAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGC 1202
Db     153429 GCTCGAGCCCGACGGCGACGCGCCCCCGCTGGTGCTGCGCGACGACGCGGACGCGGGCCC 153488

Qy       1203 GCACACCC 1210
Db     153489 GCCCCCGC 153496


RESULT 12
US-09-827-688-8/c
; Sequence 8, Application US/09827688
; Patent No. 6821955
; GENERAL INFORMATION:
;  APPLICANT: ORSON, FRANK
;  APPLICANT:  KINSEY, BERMA
;  APPLICANT:  BHOGAL, BALBIR
;  TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION DE
;  TITLE OF INVENTION:  AGENTS
;  FILE REFERENCE: P01949US1/10004014
;  CURRENT APPLICATION NUMBER: US/09/827,688
;  CURRENT FILING DATE:  2001-04-06
;  PRIOR APPLICATION NUMBER: 60/195,680
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 8
;   LENGTH: 154746
;   TYPE: DNA
;   ORGANISM: HERPESVIRUS 2
US-09-827-688-8

  Query Match             6.6%;  Score 82.4;  DB 3;  Length 154746;
  Best Local Similarity   44.6%;  Pred. No. 7.9e-06;
  Matches  458;  Conservative    0;  Mismatches  561;  Indels    9;  Gaps    3;

Qy        183 CGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCAT 242
Db     129264 CGGCGGCGCGCCCGCCGCCTTCGGCCCGCTGCGCGCCTCGGGCCCGCTGCGCCGCGCGGC 129205

Qy        243 CGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTC 302
Db     129204 GGCCTGGATGCGCCAGGTGCCCGACCCGGAGGACGTGCGCGTGGTGATCCTCTACTCGCC 129145

Qy        303 CCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCT 362
Db     129144 GCTGCCGGGCGAGGACCTGGCCGCGGGCCGCGCCGGGGGCGGGCCCCCCCCGGAGTGGTC 129085

Qy        363 GCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTG 422
Db     129084 CGCCGAGCGCGGCGGGCTGTCCTGCCTG---CTGGCGGCCCTGGGCAACCGGCTCTGCGG 129028

Qy        423 GTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCG 482
Db     129027 GCCCGCCACGGCCGCCTGGGCGGGCAACTGGACCGGCGCCCCCGACGTCTCGGCGCTGGG 128968

Qy        483 CAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTG 542
Db     128967 CGCGCAGGGCGTGCTGCTGCTGTCCACGCGGGACCTGGCCTTCGCCGGCGCCGTGGAGTT 128908

Qy        543 TGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCT 602
Db     128907 CCTGGGGCTGCTGGCCGGCGCCTGCGACCGCCGCCTCATCGTCGTCAACGCCGT---GCG 128851

Qy        603 CAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGG 662
Db     128850 CGCCGCGGACTGGCCCGCCGACGGGCCCGTGGTCTCGCGGCAGCACGCCTACCTGGCCTG 128791

Qy        663 CATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTG 722
Db     128790 CGAGGTGCTGCCCGCCGTGCAGTGCGCCGTGCGCTGGCCGGCGGCGCGGGACCTGCGCCG 128731

Qy        723 CACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCT 782
Db     128730 CACCGTGCTGGCCTCCGGCCGCGTGTTCGGGCCGGGGGTCTTCGCGCGCGTGGAGGCCGC 128671

Qy        783 CGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGT 842
Db     128670 GCACGCGCGCCTGTACCCCGACGCGCCGCCGCTGCGCCTCTGCCGCGGGGCCAACGTGCG 128611

Qy        843 GCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTT 902
Db     128610 GTACCGCGTGCGCACGCGCTTCGGCCCCGACACGCTGGTGCCCATGTCCCCGCGCGAGTA 128551

Qy        903 CGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGA 962
Db     128550 CCGCCGCGCCGTGCTCCCGGCGCTGGACGGCCGGGCCGCCGCCTCGGGCGCGGGCGACGC 128491

Qy        963 TCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCT 1022
Db     128490 CATGGCGCCCGGCGCGCCGGACTTCTGCGAGGACGAG---GCGCACTCGCACCGCGCCTG 128434

Qy       1023 GGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCG 1082
Db     128433 CGCGCGCTGGGGCCTGGGCGCGCCGCTGCGGCCCGTCTACGTGGCGCTGGGGCGCGACGC 128374

Qy       1083 CCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCG 1142
Db     128373 CGTGCGCGGCGGCCCGGCGGAGCTGCGCGGGCCGCGGCGGGAGTTCTGCGCGCGGGCGCT 128314

Qy       1143 GAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGC 1202
Db     128313 GCTCGAGCCCGACGGCGACGCGCCCCCGCTGGTGCTGCGCGACGACGCGGACGCGGGCCC 128254

Qy       1203 GCACACCC 1210
Db     128253 GCCCCCGC 128246


RESULT 13
US-09-758-759-46
; Sequence 46, Application US/09758759
; Patent No. 6861513
; GENERAL INFORMATION:
;  APPLICANT: Hosted, Thomas J.
;  APPLICANT:  Wang, Tim X.
;  APPLICANT:  Horan, Ann C.
;  TITLE OF INVENTION: Everninomicin Biosynthetic Genes
;  FILE REFERENCE: ID0983K US
;  CURRENT APPLICATION NUMBER: US/09/758,759
;  CURRENT FILING DATE:  2001-01-11
;  PRIOR APPLICATION NUMBER: US 60/175,751
;  PRIOR FILING DATE: 2000-01-12
;  NUMBER OF SEQ ID NOS: 204
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
;   LENGTH: 1317
;   TYPE: DNA
;   ORGANISM: Micromonospora carbonacea
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1317)
;   OTHER INFORMATION: evrK
US-09-758-759-46

  Query Match             5.3%;  Score 66.6;  DB 3;  Length 1317;
  Best Local Similarity   44.7%;  Pred. No. 0.0042;
  Matches  301;  Conservative    0;  Mismatches  369;  Indels    3;  Gaps    1;

Qy        513 CGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGC 572
Db        264 CGTGGCCGGTCTGCACCTGGACCTGCGGCTGGTCCGGCGGCGGGCCGGCACGATCGGCGC 323

Qy        573 GAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCC 632
Db        324 GGTGGCCGTCGGCGGCCTCGCGGTGCCCCTCGGCCTGGGCATCGCCGCCGGCCTGCTGGC 383

Qy        633 GCCGGTGCGCATGCTGGAGTTCGTCCTCG---GCATCGTGATGGCGCTGATCCTGCGCGC 689
Db        384 CCCGGCGGCGCTTCTCGCGGCCGGGCAGGAGCGGACTGTCTTCGCGCTGTTCGTCGGCGT 443

Qy        690 GGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGG 749
Db        444 GGCGATGGCGGTCAGCGCCGTGCCGGTGATCGCGAAGACGCTCACCGACATGCGCCTGCT 503

Qy        750 CCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGC 809
Db        504 GCACCGCGACGTGGGGCAGATCATCCTGGCTGCGGCGTCGCTGGAGGACGCTGCGGCCTG 563

Qy        810 GCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTC 869
Db        564 GTTCCTGCTGTCGCTCATCTCGTCGGTGGCGGTGAGCACCCTCACCGCCGGGCAGGTGGT 623

Qy        870 GGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGAT 929
Db        624 GACCGCCCTGCTTTACCTCGTGGCCTACCTCGCGGTGGCCGTCCTGGTCGGCCGGCCGGT 683

Qy        930 CATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCAC 989
Db        684 GACCCGGCGCGCCATGCGCTGGGCGAACGCCCAGCCCGACGGCGGGGCCGCCAGCGCCGT 743

Qy        990 CCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCT 1049
```

```
Db        744 CGCCGTGGTGATCGTGCTGGCCTTCGCGGCGGGGGGCGCACGCGCTGGGCCTGGAGGCGAT 803

Qy       1050 GCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTC 1109

Db        804 CTTCGGCGCGCTGGTGGCGGGTGTCCTGATCGGCCTTCCCGGCAACGGGGAGCCGGCCCG 863

Qy       1110 CCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGG 1169

Db        864 GCTGGCACCGCTGCGCACGGTGGTGCTGTCCGTGCTCGCCCCGATCTTCCTGGCCAGCGC 923

Qy       1170 CGTGCCCCTGCTC 1182

Db        924 GGGGCTCCGGGTC 936


RESULT 14
US-09-758-759-1/c
; Sequence 1, Application US/09758759
; Patent No. 6861513
; GENERAL INFORMATION:
;  APPLICANT: Hosted, Thomas J.
;  APPLICANT:  Wang, Tim X.
;  APPLICANT:  Horan, Ann C.
;  TITLE OF INVENTION: Everninomicin Biosynthetic Genes
;  FILE REFERENCE: ID0983K US
;  CURRENT APPLICATION NUMBER: US/09/758,759
;  CURRENT FILING DATE:  2001-01-11
;  PRIOR APPLICATION NUMBER: US 60/175,751
;  PRIOR FILING DATE: 2000-01-12
;  NUMBER OF SEQ ID NOS: 204
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 109519
;   TYPE: DNA
;   ORGANISM: Micromonospora carbonacea
US-09-758-759-1

  Query Match             5.3%;  Score 66.6;  DB 3;  Length 109519;
  Best Local Similarity   44.7%;  Pred. No. 0.0044;
  Matches  301;  Conservative    0;  Mismatches  369;  Indels    3;  Gaps    1;

Qy        513 CGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGC 572

Db      31613 CGTGGCCGGTCTGCACCTGGACCTGCGGCTGGTCCGGCGGCGGGCCGGCACGATCGGCGC 31554

Qy        573 GAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCC 632

Db      31553 GGTGGCCGTCGGCGGCCTCGCGGTGCCCCTCGGCCTGGGCATCGCCGCCGGCCTGCTGGC 31494

Qy        633 GCCGGTGCGCATGCTGGAGTTCGTCCTCG---GCATCGTGATGGCGCTGATCCTGCGCGC 689

Db      31493 CCCGGCGGCGCTTCTCGCGGCCGGGCAGGAGCGGACTGTCTTCGCGCTGTTCGTCGGCGT 31434

Qy        690 GGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGG 749

Db      31433 GGCGATGGCGGTCAGCGCCGTGCCGGTGATCGCGAAGACGCTCACCGACATGCGCCTGCT 31374

Qy        750 CCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGC 809

Db      31373 GCACCGCGACGTGGGGCAGATCATCCTGGCTGCGGCGTCGCTGGAGGACGCTGCGGCCTG 31314

Qy        810 GCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTC 869

Db      31313 GTTCCTGCTGTCGCTCATCTCGTCGGTGGCGGTGAGCACCCTCACCGCCGGGCAGGTGGT 31254

Qy        870 GGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGAT 929

Db      31253 GACCGCCCTGCTTTACCTCGTGGCCTACCTCGCGGTGGCCGTCCTGGTCGGCCGGCCGGT 31194

Qy        930 CATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCAC 989

Db      31193 GACCCGGCGCGCCATGCGCTGGGCGAACGCCCAGCCCGACGGCGGGGCCGCCAGCGCCGT 31134

Qy        990 CCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCT 1049

Db      31133 CGCCGTGGTGATCGTGCTGGCCTTCGCGGCGGGGGCGCACGCGCTGGGCCTGGAGGCGAT 31074

Qy       1050 GCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTC 1109

Db      31073 CTTCGGCGCGCTGGTGGCGGGTGTCCTGATCGGCCTTCCCGGCAACGGGGAGCCGGCCCG 31014

Qy       1110 CCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGG 1169

Db      31013 GCTGGCACCGCTGCGCACGGTGGTGCTGTCCGTGCTCGCCCCGATCTTCCTGGCCAGCGC 30954

Qy       1170 CGTGCCCCTGCTC 1182

Db      30953 GGGGCTCCGGGTC 30941


RESULT 15
US-08-690-473-1
; Sequence 1, Application US/08690473
; Patent No. 5876923
;  GENERAL INFORMATION:
;    APPLICANT:  Leopardi, Rosario
;    APPLICANT:  Roizman, Bernard
;    TITLE OF INVENTION:  HERPES SIMPLEX VIRUS ICP4 AS AN
;    TITLE OF INVENTION:  INHIBITOR OF APOPTOSIS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/690,473
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  ARCD:239
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4257 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-690-473-1

  Query Match             5.2%;  Score 64.8;  DB 2;  Length 4257;
  Best Local Similarity   42.8%;  Pred. No. 0.0088;
  Matches  440;  Conservative    0;  Mismatches  582;  Indels    6;  Gaps    2;

Qy        198 GGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCC 257

Db       3057 GGGCCGCCTCTTTGGCCCCCTGCGCGCCTCGGGCCCGCTGCGCCGCATGGCGGCCTGGAT 3116

Qy        258 GCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCGGT 317

Db       3117 GCGCCAGATCCCCGACCCCGAGGACGTGCGCGTGGTGGTGCTGTACTCGCCGCTGCCGGG 3176

Qy        318 ACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCCTG 377
```

```
Db       3177 CGAGGACCTGGCCGGCGGCGGGGCCTCGGGGGGGCCGCCGGAGTGGTCCGCCGAGCGCGG 3236

Qy        378 GCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGTGA 437

Db       3237 CGGGCTGTCCTGCCTGCTGGCGGCCCTGGCCAACCGGCTGTGCGGGCCGGACACGGCCGC 3296

Qy        438 GATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACG 497

Db       3297 CTGGGCGGGCAATTGGACCGGCGCCCCCGACGTGTCGGCGCTGGGCGCACAGGGCGTGCT 3356

Qy        498 GCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGC 557

Db       3357 GCTGCTGTCCACGCGGGACCTGGCCTTCGCCGGGGCCGTGGAGTTTCTGGGGCTGCTCGC 3416

Qy        558 GGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTT 617

Db       3417 CAGCGCCGGCGACCGGCGGCTCATCGTGGTCAAC---ACCGTGCGCGCCTGCGACTGGCC 3473

Qy        618 CGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCT 677

Db       3474 CGCCGACGGGCCCGCGGTGTCGCGGCAGCACGCCTACCTGGCGTGCGAGCTGCTGCCCGC 3533

Qy        678 GATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGC 737

Db       3534 CGTGCAGTGCGCCGTGCGCTGGCCGGCGGCGCGGGACCTGCGCCGCACGGTGCTGGCCTC 3593

Qy        738 CGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGT 797

Db       3594 GGGCCGCGTGTTCGGCCCGGGGGTCTTCGCGCGCGTGGAGGCCGCGCACGCGCGCCTGTA 3653

Qy        798 CGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCAC 857

Db       3654 CCCCGACGCGCCGCCGCTGCGCCTGTGCCGCGGCGGCAACGTGCGCTACCGCGTGCGCAC 3713

Qy        858 GGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGT 917

Db       3714 GCGCTTCGGCCCGGACACGCCGGTGCCCATGTCCCCGCGCGAGTACCGCCGGGCCGTGCT 3773

Qy        918 CCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCG 977

Db       3774 GCCGGCGCTGGACGGCCGGGCGGCGGCCTCGGGGACCACCGACGCCATGGCGCCCGGCGC 3833

Qy        978 GCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCT 1037

Db       3834 GCCGGACTTCTGCGAGGAGGAG---GCCCACTCGCACGCCGCCTGCGCGCGCTGGGGCCT 3890

Qy       1038 GGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCG 1097

Db       3891 GGGCGCGCCGCTGCGGCCCGTGTACGTGGCGCTGGGGCGCGAGGCGGTGCGCGCCGGCCC 3950

Qy       1098 CAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGC 1157

Db       3951 GGCCCGGTGGCGCGGGCCGCGGAGGGACTTTTGCGCCCGCGCCCTGCTGGAGCCCGACGA 4010

Qy       1158 CGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCCGCGCACACCCCCGCAGC 1217

Db       4011 CGACGCCCCCCCGCTGGTGCTGCGCGGCGACGACGACGGCCCGGGGGCCCTGCCGCCGGC 4070

Qy       1218 GACGAACG 1225

Db       4071 GCCGCCCG 4078
```

Search completed: August 25, 2006, 22:12:39
Job time : 303 secs

This Page Blank (uspto)

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2006, 21:58:01 ; Search time 2212 Seconds
(without alignments)
6915.961 Million cell updates/sec

Title: US-10-680-860A-13
Perfect score: 1245
Sequence: 1 gtgctccagcgcgtcgatct.....aacccaccccgagaggatga 1245

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_PUBCOMB.seq:*
3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
9: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
10: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
11: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
12: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
13: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
14: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11B_PUBCOMB.seq:*
15: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11C_PUBCOMB.seq:*
16: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11D_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1245	100.0	1245	10	US-10-680-860A-13	Sequence 13, Appl
	2	1245	100.0	11171	10	US-10-680-860A-2	Sequence 2, Appli
c	3	794	63.8	84428	8	US-10-229-148B-1	Sequence 1, Appli
	4	356.4	28.6	84428	8	US-10-229-148B-1	Sequence 1, Appli
c	5	251.4	20.2	19016	10	US-10-469-442-1	Sequence 1, Appli
	6	246.6	19.8	1371	7	US-10-402-842-17	Sequence 17, Appl
	7	246.6	19.8	1371	8	US-10-746-795-17	Sequence 17, Appl
	8	246.6	19.8	47988	7	US-10-402-842-1	Sequence 1, Appli
	9	246.6	19.8	48200	8	US-10-746-795-1	Sequence 1, Appli
	10	227	18.2	1227	7	US-10-402-842-16	Sequence 16, Appl
	11	227	18.2	1227	8	US-10-746-795-16	Sequence 16, Appl
c	12	215.6	17.3	9025608	7	US-10-156-761-1	Sequence 1, Appli
	13	215	17.3	1176	7	US-10-156-761-3774	Sequence 3774, Ap
	14	184	14.8	1137	12	US-10-496-351-23	Sequence 23, Appl
c	15	184	14.8	82746	12	US-10-496-351-56	Sequence 56, Appl
	16	86.6	7.0	1826	6	US-10-138-632-14	Sequence 14, Appl
	17	86.6	7.0	1826	9	US-10-864-423-14	Sequence 14, Appl
	18	82.4	6.6	3957	7	US-10-200-562-193	Sequence 193, App
	19	82.4	6.6	3957	7	US-10-237-551-193	Sequence 193, App
	20	82.4	6.6	3957	10	US-10-945-050-193	Sequence 193, App
	21	82.4	6.6	154746	3	US-09-827-688-8	Sequence 8, Appli
c	22	82.4	6.6	154746	3	US-09-827-688-8	Sequence 8, Appli
	23	73.4	5.9	1266	7	US-10-156-761-4878	Sequence 4878, Ap
	24	72.2	5.8	1431	9	US-10-425-115-25358	Sequence 25358, A
	25	70.6	5.7	1431	8	US-10-425-114-33423	Sequence 33423, A
	26	69.8	5.6	125401	8	US-10-203-295-35	Sequence 35, Appl
	27	68.6	5.5	1341	7	US-10-156-761-2330	Sequence 2330, Ap
	28	68.2	5.5	14055	12	US-10-496-351-1	Sequence 1, Appli
	29	67.2	5.4	1458	7	US-10-156-761-1306	Sequence 1306, Ap
	30	66.6	5.3	1317	3	US-09-758-759-46	Sequence 46, Appl
	31	66.6	5.3	1317	13	US-11-021-825-46	Sequence 46, Appl
	32	66.6	5.3	1338	7	US-10-107-431-92	Sequence 92, Appl
	33	66.6	5.3	15240	7	US-10-107-431-278	Sequence 278, App
c	34	66.6	5.3	109519	3	US-09-758-759-1	Sequence 1, Appli
c	35	66.6	5.3	109519	13	US-11-021-825-1	Sequence 1, Appli
c	36	65.8	5.3	1381	8	US-10-374-780A-1512	Sequence 1512, Ap
c	37	65.8	5.3	1381	8	US-10-412-699B-1592	Sequence 1592, Ap
	38	65.6	5.3	2115	7	US-10-156-761-5640	Sequence 5640, Ap
	39	65.6	5.3	9025608	7	US-10-156-761-1	Sequence 1, Appli
c	40	65.2	5.2	1869	9	US-10-211-028-22	Sequence 22, Appl
	41	65.2	5.2	2154	7	US-10-156-761-6857	Sequence 6857, Ap
c	42	65.2	5.2	90597	9	US-10-211-028-1	Sequence 1, Appli
	43	65	5.2	1650	8	US-10-282-122A-25691	Sequence 25691, A
	44	64.8	5.2	4257	3	US-09-825-288A-1	Sequence 1, Appli
c	45	64.8	5.2	135638	7	US-10-314-657-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-680-860A-13
; Sequence 13, Application US/10680860A
; Publication No. US20050202528A1
; GENERAL INFORMATION:
; APPLICANT: BLONDELET-ROUAULT, Marie-Helene
; APPLICANT: DOMINGUEZ, Helene
; APPLICANT: DARBON-RONGERE, Emmanuelle
; APPLICANT: GERBAUD, Claude
; APPLICANT: GONDRAN, Anne
; APPLICANT: KARRAY, Fatma
; APPLICANT: LACROIX, Patricia
; APPLICANT: OESTREICHER-MERMET-BOUVIER, Nathalie
; APPLICANT: PERNODET, Jean-Luc
; APPLICANT: TUPHILE, Karine
; TITLE OF INVENTION: POLYPEPTIDES INVOLVED IN THE BIOSYNTHESIS OF SPIRAMYCINS, NUCLEOTI
; TITLE OF INVENTION: SEQUENCE ENCODING THESE POLYPEPTIDES AND APPLICATIONS THEREOF
; FILE REFERENCE: FRAV2002/0028 US NP
; CURRENT APPLICATION NUMBER: US/10/680,860A
; 2003-10-07
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: FR 0212489
; PRIOR FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: FR 0302439
; PRIOR FILING DATE: 2003-02-27
; PRIOR APPLICATION NUMBER: US 60/493,490
; PRIOR FILING DATE: 2003-08-07
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Streptomyces ambofaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1245)
US-10-680-860A-13

Query Match 100.0%; Score 1245; DB 10; Length 1245;
Best Local Similarity 100.0%; Pred. No. 3.4e-293;

```
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60

Qy         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120

Qy        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180

Qy        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240

Qy        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300

Qy        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360

Qy        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420

Qy        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480

Qy        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540

Qy        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600

Qy        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660

Qy        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720

Qy        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780

Qy        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840

Qy        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900

Qy        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960

Qy        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020

Qy       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080

Qy       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140

Qy       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200

Qy       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245


RESULT 2
US-10-680-860A-2
; Sequence 2, Application US/10680860A
; Publication No. US20050202528A1
; GENERAL INFORMATION:
;  APPLICANT: BLONDELET-ROUAULT, Marie-Helene
;  APPLICANT:  DOMINGUEZ, Helene
;  APPLICANT:  DARBON-RONGERE, Emmanuelle
;  APPLICANT:  GERBAUD, Claude
;  APPLICANT:  GONDRAN, Anne
;  APPLICANT:  KARRAY, Fatma
;  APPLICANT:  LACROIX, Patricia
;  APPLICANT:  OESTREICHER-MERMET-BOUVIER, Nathalie
;  APPLICANT:  PERNODET, Jean-Luc
;  APPLICANT:  TUPHILE, Karine
;  TITLE OF INVENTION: POLYPEPTIDES INVOLVED IN THE BIOSYNTHESIS OF SPIRAMYCINS, NUCLEOTI
;  TITLE OF INVENTION:  SEQUENCE ENCODING THESE POLYPEPTIDES AND APPLICATIONS THEREOF
;  FILE REFERENCE: FRAV2002/0028 US NP
;  CURRENT APPLICATION NUMBER: US/10/680,860A
;    2003-10-07
;  CURRENT FILING DATE:  2003-10-07
;  PRIOR APPLICATION NUMBER: FR 0212489
;  PRIOR FILING DATE: 2002-10-08
;  PRIOR APPLICATION NUMBER: FR 0302439
;  PRIOR FILING DATE: 2003-02-27
;  PRIOR APPLICATION NUMBER: US 60/493,490
;  PRIOR FILING DATE: 2003-08-07
;  NUMBER OF SEQ ID NOS: 161
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 11171
;   TYPE: DNA
;   ORGANISM: Streptomyces ambofaciens
US-10-680-860A-2

  Query Match            100.0%;  Score 1245;  DB 10;  Length 11171;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-293;
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4061 GTGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACG 4120

Qy         61 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4121 GTATTCGCCTGTCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCA 4180

Qy        121 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4181 CTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGA 4240

Qy        181 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4241 TTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGC 4300

Qy        241 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4301 ATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTC 4360
```

```
Qy        301 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4361 TCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTT 4420

Qy        361 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4421 CTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGC 4480

Qy        421 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4481 TGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTG 4540

Qy        481 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4541 CGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACG 4600

Qy        541 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4601 TGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCG 4660

Qy        601 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4661 CTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTC 4720

Qy        661 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4721 GGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTC 4780

Qy        721 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4781 TGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACC 4840

Qy        781 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4841 CTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGAC 4900

Qy        841 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4901 GTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCC 4960

Qy        901 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4961 TTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGC 5020

Qy        961 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5021 GATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATG 5080

Qy       1021 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5081 CTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATG 5140

Qy       1081 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5141 CGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCC 5200

Qy       1141 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5201 CGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGCGTACCCGGGCCC 5260

Qy       1201 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       5261 GCGCACACCCCCGCAGCGACGAACGAACCCACCCCGAGAGGATGA 5305


RESULT 3
US-10-229-148B-1/c
; Sequence 1, Application US/10229148B
; Publication No. US20040091975A1
; GENERAL INFORMATION:
;  APPLICANT: Meiji Seika Kaisha, Ltd.
;  TITLE OF INVENTION: Midecamycin biosynthetic genes
;  FILE REFERENCE: 138451 US
;  CURRENT APPLICATION NUMBER: US/10/229,148B
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 210516/2002
;  PRIOR FILING DATE: 2002-07-19
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 84428
;   TYPE: DNA
;   ORGANISM: Streptomyces mycarofaciens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((1)..(675))
;   OTHER INFORMATION: ORF42 (fragment)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((1168)..(2202))
;   OTHER INFORMATION: ORF41
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((2220)..(3215))
;   OTHER INFORMATION: ORF40
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((3237)..(4691))
;   OTHER INFORMATION: ORF39
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((4695)..(5948))
;   OTHER INFORMATION: ORF38
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((6048)..(6629))
;   OTHER INFORMATION: ORF37
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((6653)..(7945))
;   OTHER INFORMATION: ORF36
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (8149)..(9015)
;   OTHER INFORMATION: ORF35
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (9012)..(9335)
;   OTHER INFORMATION: ORF34
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (9328)..(10458)
;   OTHER INFORMATION: ORF33
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10521)..(11603)
;   OTHER INFORMATION: ORF32
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (11729)..(12961)
;   OTHER INFORMATION: ORF31
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((13016)..(14044))
;   OTHER INFORMATION: ORF30
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((14074)..(15096))
;   OTHER INFORMATION: ORF29
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (15643)..(17466)
;   OTHER INFORMATION: ORF28
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: Complement((17522)..(18895))
```

; OTHER INFORMATION: ORF27
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19063)..(20229)
; OTHER INFORMATION: ORF26
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((20307)..(21743))
; OTHER INFORMATION: ORF25
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((21733)..(22527))
; OTHER INFORMATION: ORF24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((22534)..(23571))
; OTHER INFORMATION: ORF23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((23555)..(24463))
; OTHER INFORMATION: ORF22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((24460)..(25650))
; OTHER INFORMATION: ORF21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((25647)..(26105))
; OTHER INFORMATION: ORF20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((26180)..(27391))
; OTHER INFORMATION: ORF19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27937)..(28983)
; OTHER INFORMATION: ORF18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29244)..(42779)
; OTHER INFORMATION: ORF1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42823)..(48657)
; OTHER INFORMATION: ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (48712)..(59802)
; OTHER INFORMATION: ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59850)..(64556)
; OTHER INFORMATION: ORF4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64687)..(70365)
; OTHER INFORMATION: ORF5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70365)..(71078)
; OTHER INFORMATION: ORF6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71113)..(72360)
; OTHER INFORMATION: ORF7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72400)..(73665)
; OTHER INFORMATION: ORF8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73694)..(75043)
; OTHER INFORMATION: ORF9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((75899)..(76570))
; OTHER INFORMATION: ORF10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((76602)..(77765))
; OTHER INFORMATION: ORF11
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78039)..(79313)
; OTHER INFORMATION: ORF12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((79391)..(81052))
; OTHER INFORMATION: ORF13
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81541)..(82356)
; OTHER INFORMATION: ORF14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82760)..(83362)
; OTHER INFORMATION: ORF15
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((83495)..(84142))
; OTHER INFORMATION: ORF16
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (84329)..(84428)
; OTHER INFORMATION: ORF17 (fragment)
US-10-229-148B-1

Query Match 63.8%; Score 794; DB 8; Length 84428;
Best Local Similarity 80.9%; Pred. No. 1.3e-183;
Matches 937; Conservative 0; Mismatches 220; Indels 1; Gaps 1;

```
Qy        12 CGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTG 71
             ||||   ||| |||| |||||||||||| |||||||  |||||  |  ||||||||||||
Db     77751 CGTCCGCCTGCCGTCGCTCACCGGCCTTCGCTGGTTCGCGGCCTTAGCGGTATTCGCCTG 77692

Qy        72 TCACATCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACAT 131
             |||||| |||||||||   |||||||| |||||||||| || | ||| ||||| |||||
Db     77691 TCACATTGCCCAGCAGCAATTCTTCGCTGACCAGCAGGTCGGAACCGCGCTGCTACACAT 77632

Qy       132 CACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGC 191
             |||| |||||||||| || |||||||| |||||||| | || || || || || || ||
Db     77631 CACCACGCTCGGTTCGATCGCGGTCTCGGTCTTCTTCCTGCTCAGCGGTTTTGTTCTGGC 77572

Qy       192 CTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGAT 251
             |||||||||||  ||| ||||||| ||  ||||||||||||||||||  | || |||||
Db     77571 GTGGTCGGCCCGTGACAAGGACTCTGTCACGACTTTCTGGCGGCGCCGATTTGCCAAGAT 77512

Qy       252 CTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGA 311
             ||| |||||||| |||   || |||  ||| ||||    |||||||||||||||   ||||
Db     77511 CTACCCGCTGCACCTCGTAACTTTCCTCATAGCGGGCGTCATCATTTTCTCCCTCGCGGA 77452

Qy       312 GCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCA 371
             ||||   || ||||| |||||||||||||||| ||||| ||| || ||||||| |||||
Db     77451 GCCGACTCTGCCCGGTGGTTCCGTATGGGACGGTCTGGTTCCCGATCTTCTGCTGGTGCA 77392

Qy       372 GTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTC 431
             |||||||||||||| | ||  | |  ||| | |||||||| || || ||||||||||
Db     77391 GTCCTGGCTTCCCGAACCCACGATTATCGCCGGGTTCAACACCCCGAGTTGGTCGCTCTC 77332

Qy       432 CTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCC 491
             ||||||  |||| |||||||||| |||||||||||||||| |||||||||| | ||||||
Db     77331 CTGTGAATTCGCTTTCTATCTGACGTTCCCGCTGTGGTATCGGCTGGTGCGGAAGATTCC 77272

Qy       492 CGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCT 551
             | |||||||||||||||| ||||| || || |||| |||||||  ||||| ||  |
Db     77271 GGTACGGCGGCTGTGGTGGTGTGCCGCCGGTATTGCCGCGGCCGTGATTTGTGTACCGTT 77212
```

```
Qy        552 GCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGT 611
              || || ||| | |||||||||||||  |||  |||||||||||||| ||||||||||| |
Db      77211 CGTGACGAGCCAGTTCCCGGCGAGCGCCGAGACGGCCCCCGGGATGCCGCTCAACGAGCT 77152

Qy        612 CTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGAT 671
              |||||||| | |||||| |||||||||||| ||||||||||||||||||||||||||||
Db      77151 GTGGTTCGCCTGCTGGCTGCCGCCGGTGCGGATGCTGGAGTTCGTCCTCGGCATCGTGAT 77092

Qy        672 GGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCT 731
              ||||||||||||||| ||||   | |||  |||||||||    || |  || | ||  |||
Db      77091 GGCGCTGATCCTGCGCACGGGCGTGTGGCGGGGCCCCGGAGTGGTGTCCTCCGCGCTGCT 77032

Qy        732 CCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTG 791
              |||||||||   |||||| |||||||| |||||||||||| |||||| ||| ||| ||
Db      77031 GCTCGCCGCGGCCTACGGCGTCACCCAGGTGGTGCCCCCGATGTTCACCATCGCCGCGTG 76972

Qy        792 CTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCG 851
              |||| |||| || ||||| |||||||||| ||  ||||| | ||||| ||||| ||||
Db      76971 CTCCATCGTGCCCGCCGCCCTGCTGATCACCGCTTTGGCCAATGCCGATGTGCAGGGCCT 76912

Qy        852 GCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTA 911
              ||||||  ||  |||| ||||||   |||||||||||||| ||||||||||||||||||
Db      76911 GCGCACCGGTTTGCGCTCGGCGGTACTGGTGCGGCTGGGCGAGTGGTCCTTCGCCTTCTA 76852

Qy        912 CCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTA 971
              |||||  ||||||  || ||||||||||||| |||||||||||||||||||| ||||||||
Db      76851 TCTGGTTCACTTCATGGTCATCCGCTACGGGCACCGGCTGATGGGCGGCGAACTGGGCTA 76792

Qy        972 CGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGC 1031
              || ||| |||||||||||| ||  ||||   |||||| |||||| |||||||| ||| ||
Db      76791 CGCGCGCCAGTGGAGCACCGCGAGCGCGGGAGCGCTGGCCCTGGCGATGCTGGCGGTCGC 76732

Qy       1032 GGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGG 1091
              | || ||||||| || |||||||||||||||| ||  | |||||||||| ||| ||||
Db      76731 GATCGTGGCCGGTGGGCTGCTGCACACCGTCGTGGAGAACCCCTGCATGCGTCTGCTCGG 76672

Qy       1092 CAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCG-CGGCTCCCCGGAACTCAC 1150
              | ||||| |  | | | ||    |      |  ||| |||| ||   | ||| |||||||
Db      76671 CCGCCGCCGCCCGGTCGCCACCGCACCAGACCCCGCAACCGACGAGGCTCCGAAACTCAC 76612

Qy       1151 CCGCGGCCGACGCGGCCG 1168
              ||||| | ||||   |||
Db      76611 CCGCGCCTGACGTCACCG 76594
```

RESULT 4
US-10-229-148B-1
; Sequence 1, Application US/10229148B
; Publication No. US20040091975A1
; GENERAL INFORMATION:
; APPLICANT: Meiji Seika Kaisha, Ltd.
; TITLE OF INVENTION: Midecamycin biosynthetic genes
; FILE REFERENCE: 138451 US
; CURRENT APPLICATION NUMBER: US/10/229,148B
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 210516/2002
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 84428
; TYPE: DNA
; ORGANISM: Streptomyces mycarofaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((1)..(675))
; OTHER INFORMATION: ORF42 (fragment)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((1168)..(2202))
; OTHER INFORMATION: ORF41
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((2220)..(3215))
; OTHER INFORMATION: ORF40
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((3237)..(4691))
; OTHER INFORMATION: ORF39
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((4695)..(5948))
; OTHER INFORMATION: ORF38
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6048)..(6629))
; OTHER INFORMATION: ORF37
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6653)..(7945))
; OTHER INFORMATION: ORF36
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8149)..(9015)
; OTHER INFORMATION: ORF35
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9012)..(9335)
; OTHER INFORMATION: ORF34
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9328)..(10458)
; OTHER INFORMATION: ORF33
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10521)..(11603)
; OTHER INFORMATION: ORF32
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11729)..(12961)
; OTHER INFORMATION: ORF31
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((13016)..(14044))
; OTHER INFORMATION: ORF30
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((14074)..(15096))
; OTHER INFORMATION: ORF29
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15643)..(17466)
; OTHER INFORMATION: ORF28
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((17522)..(18895))
; OTHER INFORMATION: ORF27
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19063)..(20229)
; OTHER INFORMATION: ORF26
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((20307)..(21743))
; OTHER INFORMATION: ORF25
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((21733)..(22527))
; OTHER INFORMATION: ORF24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((22534)..(23571))
; OTHER INFORMATION: ORF23
; FEATURE:
; NAME/KEY: CDS

; LOCATION: Complement((23555)..(24463))
; OTHER INFORMATION: ORF22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((24460)..(25650))
; OTHER INFORMATION: ORF21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((25647)..(26105))
; OTHER INFORMATION: ORF20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((26180)..(27391))
; OTHER INFORMATION: ORF19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27937)..(28983)
; OTHER INFORMATION: ORF18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29244)..(42779)
; OTHER INFORMATION: ORF1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42823)..(48657)
; OTHER INFORMATION: ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (48712)..(59802)
; OTHER INFORMATION: ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59850)..(64556)
; OTHER INFORMATION: ORF4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64687)..(70365)
; OTHER INFORMATION: ORF5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70365)..(71078)
; OTHER INFORMATION: ORF6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71113)..(72360)
; OTHER INFORMATION: ORF7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72400)..(73665)
; OTHER INFORMATION: ORF8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73694)..(75043)
; OTHER INFORMATION: ORF9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((75899)..(76570))
; OTHER INFORMATION: ORF10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((76602)..(77765))
; OTHER INFORMATION: ORF11
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78039)..(79313)
; OTHER INFORMATION: ORF12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((79391)..(81052))
; OTHER INFORMATION: ORF13
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81541)..(82356)
; OTHER INFORMATION: ORF14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82760)..(83362)
; OTHER INFORMATION: ORF15
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((83495)..(84142))
; OTHER INFORMATION: ORF16
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (84329)..(84428)
; OTHER INFORMATION: ORF17 (fragment)
US-10-229-148B-1

Query Match 28.6%; Score 356.4; DB 8; Length 84428;
Best Local Similarity 59.3%; Pred. No. 4.1e-77;
Matches 666; Conservative 0; Mismatches 446; Indels 12; Gaps 3;

```
Qy          2 TGCTCCAGCGCGTCGATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGG 61
              |||  |  ||      | ||| |||| |||  |||| | ||        || || |||  ||
Db      19064 TGCCCTTGCCGAAACACCTGCCGTCGCTCGGCGGCATGCGGGCCATCGCCGCACTGGTGG 19123

Qy         62 TATTCGCCTGTCACATCG---CCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCG 118
              | |||  || ||| ||||   |||||| |   |||  |  || | ||         || ||
Db      19124 TGTTCTGCTCTCATATCGCTTCCCAGCCGTTTTTCCGCAACGCCAAGATAAACTCCACCG 19183

Qy        119 CACTGCTGCACATCACCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTG 178
              ||| | | |   |   |  || ||  |  |||||||||| |||||||||| || || |
Db      19184 CACAGGTCCCGCTGGACGTCCTGGGGCCGCTGGCGGTCTCGTTCTTCTTCATGCTCAGCG 19243

Qy        179 GATTCGTCCTCGCCTGGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCC 238
              |||||||||| ||||| ||| |     |||   |||  | ||||||| ||||
Db      19244 GATTCGTCCTCACCTGGGCGGGCATGCCCGACCCGTCCAAGCCTGCCTTCTGGCGCCGCC 19303

Qy        239 GCATCGCGAAGATCTATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTT 298
              |    |   | ||||  ||||||| ||  ||    |    || ||      ||| |
Db      19304 GTTGGGTTCGGGTCTACTCGCTGCACCTGCCGGTCCTGCTGCTCACGCTGGCGATCGTGC 19363

Qy        299 TCTCCCTGTCGGAGCCGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATG 358
              | |  |||  ||| ||      |   ||||| || || |||||||  |   |   |||
Db      19364 TGTGGCTGAAGGAACC---CAATATGGGCGGGTCGGTGTGGGACGGCTTCCTCAGCAACC 19420

Qy        359 TTCTGCTCGTGCAGTCCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCA 418
              | |||||||| ||||| |||   ||||||     ||     | |||  | ||| ||
Db      19421 TGCTGCTCGTCCAGTCGTGGTGCCCCGACTACCACCAGTACGGCAGCATGAACCCGGTGG 19480

Qy        419 GCTGGTCGCTCTCCTGTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGG 478
              |||||  |||||||| |||||    |||||     | ||||||  |   || |    |
Db      19481 CGTGGTCCCTCTCCTGCGAGATGCTGTTCTACGCCGCCTTCCCGTTCCTGTTCGCCTTCT 19540

Qy        479 TGCGCAGGATTCCCGCACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGA 538
              |   || ||| |  ||   |||||||||||  || | |  ||| ||| |||| | ||
Db      19541 TCTCCAAGATGCGTGCCGAGCGGCTGTGGTCCTGGGTCCTGGGCATCTCCGTCGTCGCCG 19600

Qy        539 CGTGTGTGCCCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGC------CCCCG 592
              ||   ||||||   || ||    ||||||||  ||  |     | || |      |||
Db      19601 CGGCCGTGCCCGCCCTCGCCCTGCTGCTCCCCTCGGCCCCCACGCTGCCCTGGGACCCGA 19660

Qy        593 GGATGTCGCTCAACGAGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGT 652
              |||  ||      | |   |||||||    |||  |||||||||||||||  ||||||| |
Db      19661 ACATGCCGGAGCTCCAATACTGGTTCATCTACATGCTTCCGCCGGTGCGGCTGCTGGAAT 19720

Qy        653 TCGTCCTCGGCATCGTGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTC 712
              |||  |||||| || |||||||| |||| |  |     ||   |||||  |||||    |
Db      19721 TCGCGCTCGGCGTCCTGATGGCGCAGATCGTCAGGCGCGGCCGCTGGATCGGCCCGACCC 19780

Qy        713 CGGCGGTCTGCACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGA 772
              |||  ||| ||| ||  |||  ||||||     | ||  ||  ||        ||||  |
Db      19781 CGGGGGTGTGCGCGCTGCTGTTCGCCGGCGCGTTCGCGCTGTCCTTCGCCCTGCCGTCCT 19840

Qy        773 TCTTCACCCTCGTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCG 832
                |  | | ||| ||     | |||   | |  |||||||||| ||    | ||||| |
```

```
Db      19841 ATCTGGCTCGCGTAGCGCCGACGGTCCCGCTGATCGCGCTGCTGCTCGGCTCCCTGGCAG 19900

Qy        833 ACGCCGACGTGCACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCC 892

Db      19901 CTGGCGACATACGCGGTACCCGGTCGTGGCTGGGCACCCGGACGATGGTGCTGCTGGGTG 19960

Qy        893 AGTGGTCCTTCGCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGA 952

Db      19961 AACTCACCTTCGCCTTCTACGTCATCCACTACCTCGTCATCCAGTACGGGCACCGCTTCC 20020

Qy        953 TGGGCGGCGATCTGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCC 1012

Db      20021 TCGGCGGTGAGCTGAGCTACTACCGACAGTGGGACACCCCGGCCGCGATCGGCCTCACCG 20080

Qy       1013 TGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGC 1072

Db      20081 TTCTCGCCCTCGGGCTCAGCGTGGGCCTCGCCGCGCTCCTCCACTTCTTCGTGGAGAAGC 20140

Qy       1073 CCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCG 1116

Db      20141 CGGTCGTCCGGGCCCTCGGCCGCTCCGGCAAGGCGTCCCGCGCG 20184


RESULT 5
US-10-469-442-1/c
; Sequence 1, Application US/10469442
; Publication No. US20050089954A1
; GENERAL INFORMATION:
;  APPLICANT: PALMU, Kaisa
;  APPLICANT:  KUNNARI, Tero
;  TITLE OF INVENTION: GENE CLUSTER FOR RABELOMYCIN BIOSYNTHESIS, AND ITS USE TO GENERAT
;  TITLE OF INVENTION:  COMPOUNDS FOR DRUG SCREENING
;  FILE REFERENCE: 0933-0208P
;  CURRENT APPLICATION NUMBER: US/10/469,442
;  CURRENT FILING DATE:  2003-08-28
;  PRIOR APPLICATION NUMBER: PCT/FI02/00214
;  PRIOR FILING DATE: 2002-03-15
;  NUMBER OF SEQ ID NOS: 22
; SEQ ID NO 1
;   LENGTH: 19016
;   TYPE: DNA
;   ORGANISM: Streptomyces sp. H021
US-10-469-442-1

  Query Match             20.2%;  Score 251.4;  DB 10;  Length 19016;
  Best Local Similarity   53.8%;  Pred. No. 1.6e-51;
  Matches  659;  Conservative    0;  Mismatches  521;  Indels   45;  Gaps    5;

Qy          19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78

Db        6660 CTCCCCTCGCTCACCGGGATGCGTTTTCTCGCCGCCCTGCTGGTGTTCTCCGTGCACGGG 6601

Qy          79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCAC---ATCACC 135

Db        6600 GCCGCCGCCGGCGTCTTCAAGGACCAGGGGGTCGCAGCCGACTACTACCGCTGGTTCGGC 6541

Qy         136 CCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGG 195

Db        6540 AACGCCGGCGCGGTCGGCGTCTCCTTCTTCTTCATGCTCAGCGGCTTCGTCCTCACCTGG 6481

Qy         196 TCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTAT 255

Db        6480 TCGGTGCGTCCCGCCGACACCGTCCGCGGATTCTGGCGGCGGCGCCTGCTGAAGATCTTC 6421

Qy         256 CCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCG 315

Db        6420 CCCAACCACCTGGTGACGTTCGTCGTGGC---------------CATCGTCCTGCTGACG 6376

Qy         316 GTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCC 375

Db        6375 GTGACCAGTACCGCGGTCGCCTTCCCCGAGACGCTCGCCAACCTGTTCCTGGTGCACGCC 6316

Qy         376 TGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGT 435

Db        6315 TGGGTGCCCGACAGCGGCTACGTCGAAACGGCCAACACCGTCAGCTGGTCGCTCTCCGTG 6256

Qy         436 GAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCA 495

Db        6255 GAGCTCCTCTTCTACCTGTCCTTTCCGCTGCTCATCAAGGGCGTGAGCCGGATCGCGCCG 6196

Qy         496 CGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTG 555

Db        6195 GCCGTGCTCTGGTACGCGGCGGGTGGCGTCGTGCTGGTCATCATGCTGATGCCGCTGATC 6136

Qy         556 GCGGGCCTGCTCCCGGCGAGCGAGGAGG------TGGCCCCCGGGATGTCGCTCAACGAG 609

Db        6135 GCGCAGGTGCTGCTGCCGGACACCCGGGCTTCATGTTCATGAAGATCTCCTGGACCCAG 6076

Qy         610 GTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTG 669

Db        6075 ATCTGGTTCGTGTACGTCTTCCCCGTCGCCCGGCTCCTCGAATTCGTGCTCGGCATGCTG 6016

Qy         670 ATGGCGCTGATCCTGC------GCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGC 723

Db        6015 CTGGCGCGCATCGTGCTGTCCGGCCGGTGGATCGGGCTCGGCGTGATACCGCCGGCCCTC 5956

Qy         724 ACGGCGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTC 783

Db        5955 CTGACGATCGCCGCCTACGTGGGGGCCGTACTGATCGACCACAACCCGCTCTACAACTAC 5896

Qy         784 GTCGCCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTG 843

Db        5895 GTGGCGATCACCGTCGTCCCGCTGGCCCTGCTGATCCCGGCCGCCGCCGCCTCCGACGCC 5836

Qy         844 CACGGCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTC 903

Db        5835 AAGGGCCGCGAGAGCTTGCTGAGCAAGCGCCCCATGGTCTGGCTCGGGGAGATCTCGTTC 5776

Qy         904 GCCTTCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGAT 963

Db        5775 GCCTTCTACTGCGTGCACTATCTGATCCTGATCTACGGTCACCGGCTGTTCGGCTCGGAC 5716

Qy         964 C---------------TGGGCTACGAGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTG 1008

Db        5715 CCCAACATCTTCGGCAAGCCCTCCGGACCCGCCTGGAGCACACCCGGCGGCCTGCTCTTC 5656

Qy        1009 TCCCTGGGGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAA 1068

Db        5655 CTGGCGGCGGCCCTCGTCGTGTCCGTGCTCGCCGCGTGGGCCCTGTACGCGATCGTCGAA 5596

Qy        1069 CAGCCCTGCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCC 1128

Db        5595 CGCCCCGTCATGCGCCGCTGGGGCCGCCCGCCGGGACCCCGGCGCACCGAGCAGACGGCC 5536

Qy        1129 ACCGCGGCTCCCCGGAACTCACCCGCGGCCGACGCGGCCGGCGTGCCCCTGCTCCCGGGC 1188

Db        5535 GTCCGGCCGGAACGGGCCGCCGAAGCCCCCGCGGCCCACTGAGGACGCGGCCCCCGACCA 5476

Qy        1189 GTACCCGGGCCCGCGCACACCCCCG 1213

Db        5475 CCGACCGCACCGCCCCACACACCCG 5451


RESULT 6
US-10-402-842-17
; Sequence 17, Application US/10402842
; Publication No. US20030219872A1
; GENERAL INFORMATION:
;  APPLICANT: Magarvey, Nathan A.
;  APPLICANT:  Hucul, John A.
;  TITLE OF INVENTION: NON-RIBOSOMAL PEPTIDE SYNTHETASES AND ASSOCIATED BIOSYNTHETIC GENE
;  FILE REFERENCE: 0630/1J854-US1
;  CURRENT APPLICATION NUMBER: US/10/402,842
;  CURRENT FILING DATE:  2003-03-28
;  PRIOR APPLICATION NUMBER: US 60/368,713
;  PRIOR FILING DATE: 2002-03-30
```

```
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
;   LENGTH: 1371
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
US-10-402-842-17

  Query Match            19.8%;  Score 246.6;  DB 7;  Length 1371;
  Best Local Similarity   55.0%;  Pred. No. 2.9e-50;
  Matches  611;  Conservative    0;  Mismatches  464;  Indels   36;  Gaps    5;

Qy         17 ATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACA 76
Db        224 ACCTGAGATCCCTGACCGGTTTGCGGTTCGTGGCGATGCTGCCGGTCTTCCTCACCCATG 283

Qy         77 TCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCC 136
Db        284 CGGCGTTCGAGGGCGTCTTCAGCGACGCGAAGGTGAGCTGGGGCTTCCTCGACGCGATGG 343

Qy        137 CGCT---CGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCT 193
Db        344 GGAGCACCGGCTATATGGCCGTCTCGTTCTTCTTCGTGCTCAGCGGCTTTGTGATCACGT 403

Qy        194 GGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCT 253
Db        404 GGTCGTACCGGCCCACCGACACCGCGCGCAAGTTCTGGCGCCGGCGCTTCTTCCGGGTCT 463

Qy        254 ATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGC 313
Db        464 TCCCCAACCACGTCGTGACCTATGCGCTCGC------------CCTCGGGCTGATCGCTG 511

Qy        314 CGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGT 373
Db        512 CGGTGGGGCTGAGTGTCGGCGTACTGCCC---TCGGTCACCCAGCTCTTCCTCGTCCAGT 568

Qy        374 CCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCT 433
Db        569 CCTGGGTGCCCGACCCGGCGTTCACCGACACCGGCAACAGCGTGAGCTGGTCGCTCGCGG 628

Qy        434 GTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCG 493
Db        629 TGGATGTGGTGTTCTACGCGCTCTTCCCGGTGCTGCTCACGCTGGTGAACAAGATCAAGC 688

Qy        494 CACGGCGGCTGTGGTGGTGCGCCG------CGGGGATCGCCGTGGCCGTGACGTGTGTGC 547
Db        689 CGAATCGGCTCTGGTACTGGGTCGGTGGCTCCGTCATCGGTGTGGCCGTGGTACCGGCCA 748

Qy        548 CCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACG 607
Db        749 TCGCGCTCGCCGCGCTCCCGAGCACCCCCGAGATGCCGCTCGGCGGGGTGTCCGTCAGCC 808

Qy        608 AGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCG 667
Db        809 AGTACTGGTTCACCTACTTCTTCCCGCTCTTCCGGCTGCTGGAGTGTGTGCTCGGCATGC 868

Qy        668 TGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGG 727
Db        869 TGATGGCGCGGATCGTGCTGTCCGGCAAGTGGATACGCCTGCGGGTGCTGCCCGCCGCCG 928

Qy        728 CGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCG 787
Db        929 TCCTCGTGGTGATCGCGTACTACTTCGCCCAGCAGGTCCCGTACCTCTACCGGCTGAGTG 988

Qy        788 CCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACG 847
Db        989 CGGTGACGGTGCTGCCGGTCGCGCTGCTGACGGCGGCGGCCGCGGTGGCGGACTCCGAGG 1048

Qy        848 GCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCT 907
Db       1049 GCCGGGGCACCCTGTTCGGCAGCAAGGTCATGGTCTGGTTCGGCGAACTCTCCTTCGCCT 1108

Qy        908 TCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGG 967
Db       1109 TCTACCTGCTGCACAACCTCGTCCTGAAGTACGGCCATCTGCTGCTCGGCCACACCGAGG 1168

Qy        968 GCTACGAGCGGCAG-----------TGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGG 1015
Db       1169 AGGAGGGCGAGCTGGTGGGCCACACCTGGGGCGTGCCCGAGGGAATCGCCCTGATCGCCG 1228

Qy       1016 GGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCT 1075
Db       1229 CCGCCTTCGCGGTGTCCCTGCTGCTGGCCTGGCTGCTGCACAACGGAGTGGAGAAGCAGG 1288

Qy       1076 GCATGCGCCTGTTCGGCAGCCGCAGGTCCGC 1106
Db       1289 CGATGCGCCGCTGGTCCCGACGCAAGCCGGC 1319


RESULT 7
US-10-746-795-17
; Sequence 17, Application US/10746795
; Publication No. US20040147007A1
; GENERAL INFORMATION:
;  APPLICANT: Magarvey, Nathan A.
;  APPLICANT:  Hucul, John A.
;  TITLE OF INVENTION: NON-RIBOSOMAL PEPTIDE SYNTHETASES AND ASSOCIATED BIOSYNTHETIC GENE
;  FILE REFERENCE: 00630/100J854-US1
;  CURRENT APPLICATION NUMBER: US/10/746,795
;  CURRENT FILING DATE:  2003-12-23
;  PRIOR APPLICATION NUMBER: US 60/368,713
;  PRIOR FILING DATE: 2002-03-30
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
;   LENGTH: 1371
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
US-10-746-795-17

  Query Match            19.8%;  Score 246.6;  DB 8;  Length 1371;
  Best Local Similarity   55.0%;  Pred. No. 2.9e-50;
  Matches  611;  Conservative    0;  Mismatches  464;  Indels   36;  Gaps    5;

Qy         17 ATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACA 76
Db        224 ACCTGAGATCCCTGACCGGTTTGCGGTTCGTGGCGATGCTGCCGGTCTTCCTCACCCATG 283

Qy         77 TCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCC 136
Db        284 CGGCGTTCGAGGGCGTCTTCAGCGACGCGAAGGTGAGCTGGGGCTTCCTCGACGCGATGG 343

Qy        137 CGCT---CGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCT 193
Db        344 GGAGCACCGGCTATATGGCCGTCTCGTTCTTCTTCGTGCTCAGCGGCTTTGTGATCACGT 403

Qy        194 GGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCT 253
Db        404 GGTCGTACCGGCCCACCGACACCGCGCGCAAGTTCTGGCGCCGGCGCTTCTTCCGGGTCT 463

Qy        254 ATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGC 313
Db        464 TCCCCAACCACGTCGTGACCTATGCGCTCGC------------CCTCGGGCTGATCGCTG 511

Qy        314 CGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGT 373
Db        512 CGGTGGGGCTGAGTGTCGGCGTACTGCCC---TCGGTCACCCAGCTCTTCCTCGTCCAGT 568

Qy        374 CCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCT 433
Db        569 CCTGGGTGCCCGACCCGGCGTTCACCGACACCGGCAACAGCGTGAGCTGGTCGCTCGCGG 628

Qy        434 GTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCG 493
Db        629 TGGATGTGGTGTTCTACGCGCTCTTCCCGGTGCTGCTCACGCTGGTGAACAAGATCAAGC 688
```

```
Qy        494 CACGGCGGCTGTGGTGGTGCGCCG------CGGGGATCGCCGTGGCCGTGACGTGTGTGC 547
Db        689 CGAATCGGCTCTGGTACTGGGTCGGTGGCTCCGTCATCGGTGTGGCCGTGGTACCGGCCA 748

Qy        548 CCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACG 607
Db        749 TCGCGCTCGCCGCGCTCCCGAGCACCCCCGAGATGCCGCTCGGCGGGGTGTCCGTCAGCC 808

Qy        608 AGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCG 667
Db        809 AGTACTGGTTCACCTACTTCTTCCCGCTCTTCCGGCTGCTGGAGTGTGTGCTCGGCATGC 868

Qy        668 TGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGG 727
Db        869 TGATGGCGCGGATCGTGCTGTCCGGCAAGTGGATACGCCTGCGGGTGCTGCCCGCCGCCG 928

Qy        728 CGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCG 787
Db        929 TCCTCGTGGTGATCGCGTACTACTTCGCCCAGCAGGTCCCGTACCTCTACCGGCTGAGTG 988

Qy        788 CCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACG 847
Db        989 CGGTGACGGTGCTGCCGGTCGCGCTGCTGACGGCGGCGGCCGCGGTGGCGGACTCCGAGG 1048

Qy        848 GCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCT 907
Db       1049 GCCGGGGCACCCTGTTCGGCAGCAAGGTCATGGTCTGGTTCGGCGAACTCTCCTTCGCCT 1108

Qy        908 TCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGG 967
Db       1109 TCTACCTGCTGCACAACCTCGTCCTGAAGTACGGCCATCTGCTGCTCGGCCACACCGAGG 1168

Qy        968 GCTACGAGCGGCAG------------TGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGG 1015
Db       1169 AGGAGGGCGAGCTGGTGGGCCACACCTGGGGCGTGCCCGAGGGAATCGCCCTGATCGCCG 1228

Qy       1016 GGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCT 1075
Db       1229 CCGCCTTCGCGGTGTCCCTGCTGCTGGCCTGGCTGCTGCACAACGGAGTGGAGAAGCAGG 1288

Qy       1076 GCATGCGCCTGTTCGGCAGCCGCAGGTCCGC 1106
Db       1289 CGATGCGCCGCTGGTCCCGACGCAAGCCGGC 1319


RESULT 8
US-10-402-842-1
; Sequence 1, Application US/10402842
; Publication No. US20030219872A1
; GENERAL INFORMATION:
;  APPLICANT: Magarvey, Nathan A.
;  APPLICANT:  Hucul, John A.
;  TITLE OF INVENTION: NON-RIBOSOMAL PEPTIDE SYNTHETASES AND ASSOCIATED BIOSYNTHETIC GEN
;  FILE REFERENCE: 0630/1J854-US1
;  CURRENT APPLICATION NUMBER: US/10/402,842
;  CURRENT FILING DATE:  2003-03-28
;  PRIOR APPLICATION NUMBER: US 60/368,713
;  PRIOR FILING DATE: 2002-03-30
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 47988
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
US-10-402-842-1

  Query Match             19.8%;  Score 246.6;  DB 7;  Length 47988;
  Best Local Similarity   55.0%;  Pred. No. 2.2e-50;
  Matches  611;  Conservative    0;  Mismatches  464;  Indels   36;  Gaps    5;

Qy         17 ATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACA 76
Db      33645 ACCTGAGATCCCTGACCGGTTTGCGGTTCGTGGCGATGCTGCCGGTCTTCCTCACCCATG 33704

Qy         77 TCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCC 136
Db      33705 CGGCGTTCGAGGGCGTCTTCAGCGACGCGAAGGTGAGCTGGGGCTTCCTCGACGCGATGG 33764

Qy        137 CGCT---CGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCT 193
Db      33765 GGAGCACCGGCTATATGGCCGTCTCGTTCTTCTTCGTGCTCAGCGGCTTTGTGATCACGT 33824

Qy        194 GGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCT 253
Db      33825 GGTCGTACCGGCCCACCGACACCGCGCGCAAGTTCTGGCGCCGGCGCTTCTTCCGGGTCT 33884

Qy        254 ATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGC 313
Db      33885 TCCCCAACCACGTCGTGACCTATGCGCTCGC------------CCTCGGGCTGATCGCTG 33932

Qy        314 CGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGT 373
Db      33933 CGGTGGGGCTGAGTGTCGGCGTACTGCCC---TCGGTCACCCAGCTCTTCCTCGTCCAGT 33989

Qy        374 CCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCT 433
Db      33990 CCTGGGTGCCCGACCCGGCGTTCACCGACACCGGCAACAGCGTGAGCTGGTCGCTCGCGG 34049

Qy        434 GTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCG 493
Db      34050 TGGATGTGGTGTTCTACGCGCTCTTCCCGGTGCTGCTCACGCTGGTGAACAAGATCAAGC 34109

Qy        494 CACGGCGGCTGTGGTGGTGCGCCG------CGGGGATCGCCGTGGCCGTGACGTGTGTGC 547
Db      34110 CGAATCGGCTCTGGTACTGGGTCGGTGGCTCCGTCATCGGTGTGGCCGTGGTACCGGCCA 34169

Qy        548 CCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACG 607
Db      34170 TCGCGCTCGCCGCGCTCCCGAGCACCCCCGAGATGCCGCTCGGCGGGGTGTCCGTCAGCC 34229

Qy        608 AGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCG 667
Db      34230 AGTACTGGTTCACCTACTTCTTCCCGCTCTTCCGGCTGCTGGAGTGTGTGCTCGGCATGC 34289

Qy        668 TGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGG 727
Db      34290 TGATGGCGCGGATCGTGCTGTCCGGCAAGTGGATACGCCTGCGGGTGCTGCCCGCCGCCG 34349

Qy        728 CGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCG 787
Db      34350 TCCTCGTGGTGATCGCGTACTACTTCGCCCAGCAGGTCCCGTACCTCTACCGGCTGAGTG 34409

Qy        788 CCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACG 847
Db      34410 CGGTGACGGTGCTGCCGGTCGCGCTGCTGACGGCGGCGGCCGCGGTGGCGGACTCCGAGG 34469

Qy        848 GCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCT 907
Db      34470 GCCGGGGCACCCTGTTCGGCAGCAAGGTCATGGTCTGGTTCGGCGAACTCTCCTTCGCCT 34529

Qy        908 TCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGG 967
Db      34530 TCTACCTGCTGCACAACCTCGTCCTGAAGTACGGCCATCTGCTGCTCGGCCACACCGAGG 34589

Qy        968 GCTACGAGCGGCAG------------TGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGG 1015
Db      34590 AGGAGGGCGAGCTGGTGGGCCACACCTGGGGCGTGCCCGAGGGAATCGCCCTGATCGCCG 34649

Qy       1016 GGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCT 1075
Db      34650 CCGCCTTCGCGGTGTCCCTGCTGCTGGCCTGGCTGCTGCACAACGGAGTGGAGAAGCAGG 34709

Qy       1076 GCATGCGCCTGTTCGGCAGCCGCAGGTCCGC 1106
Db      34710 CGATGCGCCGCTGGTCCCGACGCAAGCCGGC 34740
```

RESULT 9
US-10-746-795-1
; Sequence 1, Application US/10746795
; Publication No. US20040147007A1
; GENERAL INFORMATION:
; APPLICANT: Magarvey, Nathan A.
; APPLICANT: Hucul, John A.
; TITLE OF INVENTION: NON-RIBOSOMAL PEPTIDE SYNTHETASES AND ASSOCIATED BIOSYNTHETIC GEN
; FILE REFERENCE: 00630/100J854-US1
; CURRENT APPLICATION NUMBER: US/10/746,795
; CURRENT FILING DATE: 2003-12-23
; PRIOR APPLICATION NUMBER: US 60/368,713
; PRIOR FILING DATE: 2002-03-30
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 48200
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-746-795-1

Query Match 19.8%; Score 246.6; DB 8; Length 48200;
Best Local Similarity 55.0%; Pred. No. 2.2e-50;
Matches 611; Conservative 0; Mismatches 464; Indels 36; Gaps 5;

```
Qy        17 ATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACA 76
             | |||   || || |||||  | || |     |||  |||| |||| |||  |   ||
Db     33645 ACCTGAGATCCCTGACCGGTTTGCGGTTCGTGGCGATGCTGCCGGTCTTCCTCACCCATG 33704

Qy        77 TCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCC 136
               ||     ||||| |||||  ||||  | ||||| || | |   | ||  ||   |
Db     33705 CGGCGTTCGAGGGCGTCTTCAGCGACGCGAAGGTGAGCTGGGGCTTCCTCGACGCGATGG 33764

Qy       137 CGCT---CGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCT 193
             |      ||| |  ||||| |||||| |||||||| | || || || || ||  || | |
Db     33765 GGAGCACCGGCTATATGGCCGTCTCGTTCTTCTTCGTGCTCAGCGGCTTTGTGATCACGT 33824

Qy       194 GGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCT 253
             |||||  |||    |   ||| ||| ||  |  |||||||| || ||| ||      | ||||
Db     33825 GGTCGTACCGGCCCACCGACACCGCGCGCAAGTTCTGGCGCCGGCGCTTCTTCCGGGTCT 33884

Qy       254 ATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGC 313
               ||    ||  ||| ||| |  |   ||||                ||   |||   |
Db     33885 TCCCCAACCACGTCGTGACCTATGCGCTCGC------------CCTCGGGCTGATCGCTG 33932

Qy       314 CGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGT 373
             ||||     |  | |     ||||  | |     |||  || |  |  | |||||  ||||
Db     33933 CGGTGGGGCTGAGTGTCGGCGTACTGCCC---TCGGTCACCCAGCTCTTCCTCGTCCAGT 33989

Qy       374 CCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCT 433
             |||||  | ||||||  ||  |  |||| | ||     |||||      ||||||||||| |
Db     33990 CCTGGGTGCCCGACCCGGCGTTCACCGACACCGGCAACAGCGTGAGCTGGTCGCTCGCGG 34049

Qy       434 GTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCG 493
               ||  | |  |||||    |   |||||| ||  |  |  |||||||  || |||
Db     34050 TGGATGTGGTGTTCTACGCGCTCTTCCCGGTGCTGCTCACGCTGGTGAACAAGATCAAGC 34109

Qy       494 CACGGCGGCTGTGGTGGTGCGCCG------CGGGGATCGCCGTGGCCGTGACGTGTGTGC 547
             |     ||||| ||||  ||  | ||        | |  ||||  |||||||||      |
Db     34110 CGAATCGGCTCTGGTACTGGGTCGGTGGCTCCGTCATCGGTGTGGCCGTGGTACCGGCCA 34169

Qy       548 CCCTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACG 607
               | ||| || |  || |       |  |||| |  |  |  |||| ||||  ||| |
Db     34170 TCGCGCTCGCCGCGCTCCCGAGCACCCCCGAGATGCCGCTCGGCGGGGTGTCCGTCAGCC 34229

Qy       608 AGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCG 667
             ||  ||||||| | ||||   | ||||   | ||  |||||||||  || ||||||||
Db     34230 AGTACTGGTTCACCTACTTCTTCCCGCTCTTCCGGCTGCTGGAGTGTGTGCTCGGCATGC 34289

Qy       668 TGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGG 727
             |||||||||| |||| |||   | || |  ||||   |||   |   |  |  |  | | |
‹Db    34290 TGATGGCGCGGATCGTGCTGTCCGGCAAGTGGATACGCCTGCGGGTGCTGCCCGCCGCCG 34349

Qy       728 CGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCG 787
               ||| | |          |||  | || |||||  |||  ||     |||  |   ||    |
Db     34350 TCCTCGTGGTGATCGCGTACTACTTCGCCCAGCAGGTCCCGTACCTCTACCGGCTGAGTG 34409

Qy       788 CCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACG 847
             |     | ||  | |||| ||||||||||||   |||||   ||  |  || |||    | |
Db     34410 CGGTGACGGTGCTGCCGGTCGCGCTGCTGACGGCGGCGGCCGCGGTGGCGGACTCCGAGG 34469

Qy       848 GCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCT 907
             |||||  ||||   | |  |       |    ||||  || | ||| |    ||||||||||
Db     34470 GCCGGGGCACCCTGTTCGGCAGCAAGGTCATGGTCTGGTTCGGCGAACTCTCCTTCGCCT 34529

Qy       908 TCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGG 967
             |||||||| | |||  |||  || |    ||||| || | |||| | |||  |      ||
Db     34530 TCTACCTGCTGCACAACCTCGTCCTGAAGTACGGCCATCTGCTGCTCGGCCACACCGAGG 34589

Qy       968 GCTACGAGCGGCAG------------TGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGG 1015
               | |    || |            ||| ||   || |  |  ||||| |||  |   |
Db     34590 AGGAGGGCGAGCTGGTGGGCCACACCTGGGGCGTGCCCGAGGGAATCGCCCTGATCGCCG 34649

Qy      1016 GGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCT 1075
               | | |||| |  | |||   | | | ||||||||||| ||  || ||  |||
Db     34650 CCGCCTTCGCGGTGTCCCTGCTGCTGGCCTGGCTGCTGCACAACGGAGTGGAGAAGCAGG 34709

Qy      1076 GCATGCGCCTGTTCGGCAGCCGCAGGTCCGC 1106
               |||||||  |    | | |||| | | ||
Db     34710 CGATGCGCCGCTGGTCCCGACGCAAGCCGGC 34740
```

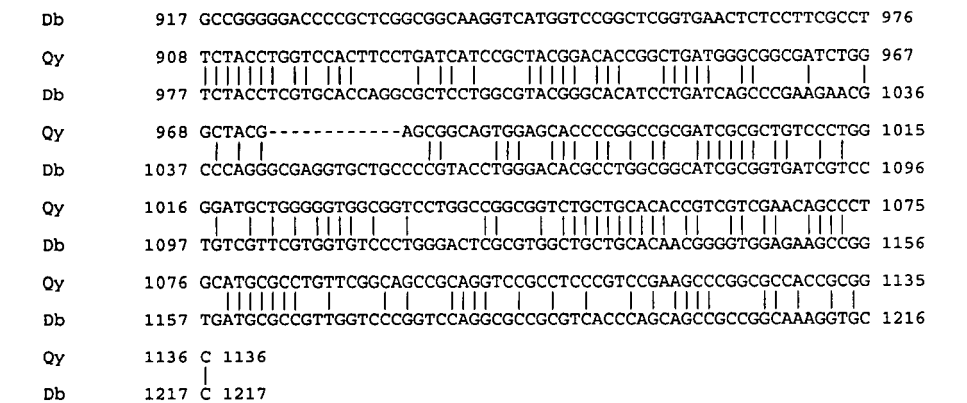
RESULT 10
US-10-402-842-16
; Sequence 16, Application US/10402842
; Publication No. US20030219872A1
; GENERAL INFORMATION:
; APPLICANT: Magarvey, Nathan A.
; APPLICANT: Hucul, John A.
; TITLE OF INVENTION: NON-RIBOSOMAL PEPTIDE SYNTHETASES AND ASSOCIATED BIOSYNTHETIC GENE
; FILE REFERENCE: 0630/1J854-US1
; CURRENT APPLICATION NUMBER: US/10/402,842
; CURRENT FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: US 60/368,713
; PRIOR FILING DATE: 2002-03-30
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-402-842-16

Query Match 18.2%; Score 227; DB 7; Length 1227;
Best Local Similarity 53.5%; Pred. No. 1.7e-45;
Matches 610; Conservative 0; Mismatches 495; Indels 36; Gaps 5;

```
Qy        17 ATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACA 76
             |||||  ||| || |||||  | || |     |       |  | || |||  |   ||
Db        92 ATCTGAGGTCCCTGACCGGATTACGGTTCCTGGGCTTATTACCCGTCTTCCTCACCCATG 151

Qy        77 TCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCAC---ATCA 133
               |||     |||||  |||||  ||||  | | ||| || | |   | ||  ||    ||
Db       152 CCGCGTTCGAGGGCGTCTTCAGCGACGCGGACGTGAGCTGGGGCTTCCTCGACGCGATGG 211

Qy       134 CCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCT 193
                 ||| |     |||||||||| |||||||| | ||||| || || ||  || |||
Db       212 GGAACACCGGCTATGCCGCGGTCTCGTTCTTCTTCGTGCTGAGCGGCTTTGTGATCACCT 271

Qy       194 GGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCT 253
             ||||   ||||  |   ||| ||   |  || |||||||| || |||  |      | |||
Db       272 GGTCCTACCGCTCCCGCGACACCACCCGCACGTTCTGGCGCCGACGCGCCTTCCGGGTCT 331

Qy       254 ATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGC 313
               ||    ||||||| | | |  |  |||||       | ||  ||  |   |
```

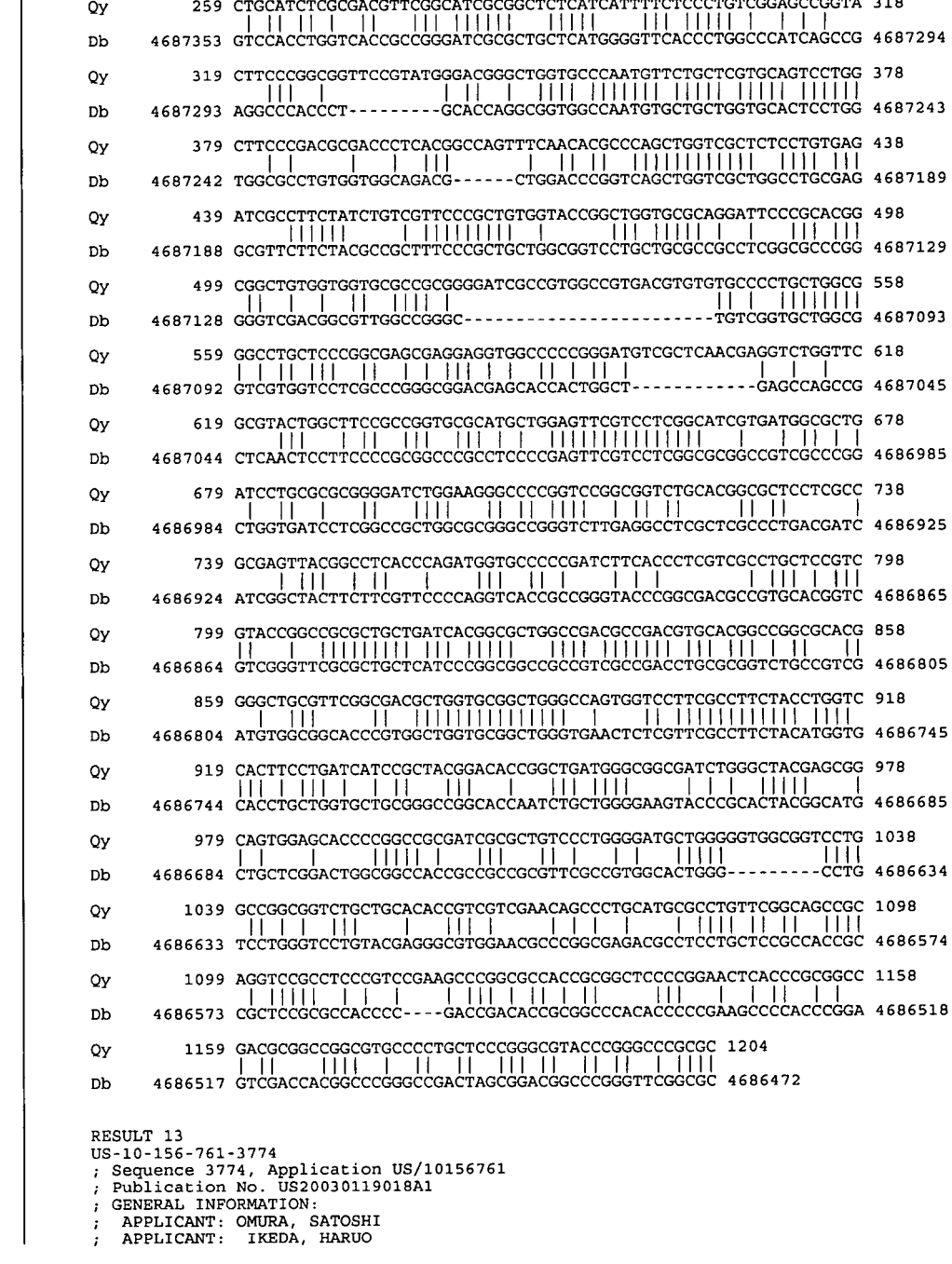
```
Db        332 TCCCCAACCATCTCGTGGCCTATGTGTTCGCGCTGGCTCTGATGCTCGCGGCGGG----- 386

Qy        314 CGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGT 373

Db        387 ----------CGCCGCCTTCGACGCCCCCGCCCTGATCTCCCAGATGTTCCTGGTGCACG 436

Qy        374 CCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCT 433

Db        437 CATGGGTGCCCGACCCGCTGTTCATCGACACCGGCAACACGGTGACCTGGTCCCTCGGGG 496

Qy        434 GTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCG 493

Db        497 TCGATGTGGTGTTCTACGGGCTCTTCCCGGTGCTGCTCGTGCTGGTGAACAAGATCAAGC 556

Qy        494 CACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGC 553

Db        557 CAACCCGTTTGTGGTACTGGGCCGGTGCTGCCGTGCTCATGGTGATCGCCATCCCCACAG 616

Qy        554 TGGCGGG---CCTGCTCCCGGCGAGCGAGGAGGT---GGCCCCCGGGATGTCGCTCAACG 607

Db        617 TGGCGCTGACCCTGCTCCCGGACACCCCGGCCATGTCGGTGGGCGATGTCTCCCGCAGCC 676

Qy        608 AGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCG 667

Db        677 AGTACTGGTTCACCTACTTCTTCCCGCTCTCCCGAACCGTGGAGTGCGTGCTGGGCATGC 736

Qy        668 TGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGG 727

Db        737 TGATGGCGCGGATCGTGCTGTCCGGCAAGTGGATAGGCCTGCGGGTGCTGCCCGCCTCGG 796

Qy        728 CGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCG 787

Db        797 CCCTGGTGGTCGTGGGGTATGTCGTCGCACAGCAACTCCCCTTCCTCTACCGGCTCAGCG 856

Qy        788 CCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACG 847

Db        857 CGGTGCTGATCGTGCCGATCGTGCTGCTCACCGCCTCCGTGGCGGTGGCCGACGCCGAGG 916

Qy        848 GCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCT 907

Db        917 GCCGGGGGACCCCGCTCGGCGGCAAGGTCATGGTCCGGCTCGGTGAACTCTCCTTCGCCT 976

Qy        908 TCTACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGG 967

Db        977 TCTACCTCGTGCACCAGGCGCTCCTGGCGTACGGGCACATCCTGATCAGCCCGAAGAACG 1036

Qy        968 GCTACG------------AGCGGCAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGG 1015

Db       1037 CCCAGGGCGAGGTGCTGCCCCGTACCTGGGACACGCCTGGCGGCATCGCGGTGATCGTCC 1096

Qy       1016 GGATGCTGGGGGTGGCGGTCCTGGCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCT 1075

Db       1097 TGTCGTTCGTGGTGTCCCTGGGACTCGCGTGGCTGCTGCACAACGGGGTGGAGAAGCCGG 1156

Qy       1076 GCATGCGCCTGTTCGGCAGCCGCAGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGG 1135

Db       1157 TGATGCGCCGTTGGTCCCGGTCCAGGCGCCGCGTCACCCAGCAGCCGCCGGCAAAGGTGC 1216

Qy       1136 C 1136

Db       1217 C 1217
```

```
RESULT 11
US-10-746-795-16
; Sequence 16, Application US/10746795
; Publication No. US20040147007A1
; GENERAL INFORMATION:
;  APPLICANT: Magarvey, Nathan A.
;  APPLICANT:  Hucul, John A.
;  TITLE OF INVENTION: NON-RIBOSOMAL PEPTIDE SYNTHETASES AND ASSOCIATED BIOSYNTHETIC GEN
;  FILE REFERENCE: 00630/100J854-US1
;  CURRENT APPLICATION NUMBER: US/10/746,795
;  CURRENT FILING DATE:  2003-12-23
;  PRIOR APPLICATION NUMBER: US 60/368,713
;  PRIOR FILING DATE: 2002-03-30
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
;   LENGTH: 1227
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
US-10-746-795-16

  Query Match             18.2%;  Score 227;  DB 8;  Length 1227;
  Best Local Similarity   53.5%;  Pred. No. 1.7e-45;
  Matches  610;  Conservative    0;  Mismatches  495;  Indels   36;  Gaps    5;

Qy         17 ATCTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACA 76

Db         92 ATCTGAGGTCCCTGACCGGATTACGGTTCCTGGGCTTATTACCCGTCTTCCTCACCCATG 151

Qy         77 TCGCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCAC---ATCA 133

Db        152 CCGCGTTCGAGGGCGTCTTCAGCGACGCGGACGTGAGCTGGGGCTTCCTCGACGCGATGG 211

Qy        134 CCCCGCTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCT 193

Db        212 GGAACACCGGCTATGCCGCGGTCTCGTTCTTCTTCGTGCTGAGCGGCTTTGTGATCACCT 271

Qy        194 GGTCGGCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCT 253

Db        272 GGTCCTACCGCTCCCGCGACACCACCCGCACGTTCTGGCGCCGACGCGCCTTCCGGGTCT 331

Qy        254 ATCCGCTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGC 313

Db        332 TCCCCAACCATCTCGTGGCCTATGTGTTCGCGCTGGCTCTGATGCTCGCGGCGGG----- 386

Qy        314 CGGTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGT 373

Db        387 ----------CGCCGCCTTCGACGCCCCCGCCCTGATCTCCCAGATGTTCCTGGTGCACG 436

Qy        374 CCTGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCT 433

Db        437 CATGGGTGCCCGACCCGCTGTTCATCGACACCGGCAACACGGTGACCTGGTCCCTCGGGG 496

Qy        434 GTGAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCG 493

Db        497 TCGATGTGGTGTTCTACGGGCTCTTCCCGGTGCTGCTCGTGCTGGTGAACAAGATCAAGC 556

Qy        494 CACGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGC 553

Db        557 CAACCCGTTTGTGGTACTGGGCCGGTGCTGCCGTGCTCATGGTGATCGCCATCCCCACAG 616

Qy        554 TGGCGGG---CCTGCTCCCGGCGAGCGAGGAGGT---GGCCCCCGGGATGTCGCTCAACG 607

Db        617 TGGCGCTGACCCTGCTCCCGGACACCCCGGCCATGTCGGTGGGCGATGTCTCCCGCAGCC 676

Qy        608 AGGTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCG 667

Db        677 AGTACTGGTTCACCTACTTCTTCCCGCTCTCCCGAACCGTGGAGTGCGTGCTGGGCATGC 736

Qy        668 TGATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGG 727

Db        737 TGATGGCGCGGATCGTGCTGTCCGGCAAGTGGATAGGCCTGCGGGTGCTGCCCGCCTCGG 796

Qy        728 CGCTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCG 787

Db        797 CCCTGGTGGTCGTGGGGTATGTCGTCGCACAGCAACTCCCCTTCCTCTACCGGCTCAGCG 856

Qy        788 CCTGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACG 847

Db        857 CGGTGCTGATCGTGCCGATCGTGCTGCTCACCGCCTCCGTGGCGGTGGCCGACGCCGAGG 916

Qy        848 GCCGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCT 907
```



RESULT 12
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 17.3%; Score 215.6; DB 7; Length 9025608;
Best Local Similarity 54.0%; Pred. No. 5.1e-43;
Matches 640; Conservative 0; Mismatches 479; Indels 67; Gaps 7;



RESULT 13
US-10-156-761-3774
; Sequence 3774, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO

```
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3774
;   LENGTH: 1176
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1176)
US-10-156-761-3774

  Query Match             17.3%;  Score 215;  DB 7;  Length 1176;
  Best Local Similarity   53.1%;  Pred. No. 1.4e-42;
  Matches  610;  Conservative    0;  Mismatches  485;  Indels   54;  Gaps    5;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
              ||| |||| |||||||||| || ||| || |||||| ||| ||||    ||  |
Db         82 CTGCCGTCCCTCACCGGCCTGCGGTGGATGGCCGCGCTGCTGGTCTTCGGACTGCATGTG 141

Qy         79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCCCG 138
                | |   |||| ||||| || |      |   ||| |||   | |    ||    |
Db        142 CGCAACTTCGGCTACTTCGACGGCACCGGTGCCCGCATCGCGACGTGGGCGTTCGGGGC- 200

Qy        139 CTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCG 198
               |||  |||  | ||| ||| ||||||||| ||    || ||||| ||    ||||||
Db        201 --CGGCGCCAGTGGGGTGTCGTTCTTCTTCATCCTCTCCGGCTTCGTGCTGATGTGGTCG 258

Qy        199 GCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCG 258
              || ||    | ||||  || ||   | |||||||||| ||||||||   ||||||||||
Db        259 GCACGGCCGGGGGACCGCGCGCTCGCCTTCTGGCGGCGGCGCATCGCCCGGATCTATCCG 318

Qy        259 CTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCTGTCGGAGCCGGTA 318
              | || || |  ||   ||| ||||||   |||||    ||| ||||| |  | | |
Db        319 GTCCACCTGGTCACCGCCGGGATCGCGCTGCTCATGGGGTTCACCCTGGCCCATCAGCCG 378

Qy        319 CTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCCTGG 378
                 |||  |          | ||  |  |||| ||||||| ||||| ||||| ||||||
Db        379 AGGCCCACC---------CTGCACCAGGCGGTGGCCAATGTGCTGCTGGTGCACTCCTGG 429

Qy        379 CTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGTGAG 438
               | |      |   |  |||       |  || ||  ||||||||||    |||| |||
Db        430 TGGCGCCTGTGGTGGCAGACG------CTGGACCCGGTCAGCTGGTCGCTGGCCTGCGAG 483

Qy        439 ATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCACGG 498
              ||||||     | |||||||||  |      ||| |||||  |  |   ||| |||
Db        484 GCGTTCTTCTACGCCGCTTTCCCGCTGCTGGCGGTCCTGCTGCGCCGCCTCGGCGCCCGG 543

Qy        499 CGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCCCCTGCTGGCG 558
               ||  |  |  ||                  |||||| |  ||  ||||    | | | |
Db        544 GGGTCGACGGCGT-----------------TGGCCGGGCTGTCGGTGCTGGCGGTCGTG 585

Qy        559 GGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAGGTCTGGTTC 618
              | |||   || ||||  |||| |       |   |   | ||||||||    ||
Db        586 GTCCTCGCCCGGGCGGACGAGCACCACTGGCTGAGCCAGCCGCTCAACTCCTTC------ 639

Qy        619 GCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTGATGGCGCTG 678
                          ||  |||  ||| | |  |||||||||||||||   |   | || | |
Db        640 -----------CCCGCGGCCCGCCTCCCCGAGTTCGTCCTCGGCGCGGCCGTCGCCCGG 687

Qy        679 ATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCGCTCCTCGCC 738
                |  ||  |  ||  ||||  || || ||||  | || ||     || ||      |

Db        688 CTGGTGATCCTCGGCCGCTGGCGCGGGCCGGGTCTTGAGGCCTCGCTCGCCCTGACGATC 747

Qy        739 GCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCCTGCTCCGTC 798
                | |||  | ||   |    |||  || |   | | |          | ||| | |||
Db        748 ATCGGCTACTTCTTCGTTCCCCAGGTCACCGCCGGGTACCCGGCGACGCCGTGCACGGTC 807

Qy        799 GTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGCCGGCGCACG 858
              ||   |  ||||||||| ||| ||||   |||| |||||||| ||| ||| | ||   ||
Db        808 GTCGGGTTCGCGCTGCTCATCCCGGCGGCCGCCGTCGCCGACCTGCGCGGTCTGCCGTCG 867

Qy        859 GGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTCTACCTGGTC 918
                |  |||      ||  |||||||||||||||  |    || ||||||||||| ||||
Db        868 ATGTGGCGGCACCCGTGGCTGGTGCGGCTGGGTGAACTCTCGTTCGCCTTCTACATGGTG 927

Qy        919 CACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGCTACGAGCGG 978
              ||| | ||| |  | ||   |||   |   ||| ||||     | | |  |||||    |
Db        928 CACCTGCTGGTGCTGCGGGCCGGCACCAATCTGCTGGGGAAGTACCCGCACTACGGCATG 987

Qy        979 CAGTGGAGCACCCCGGCCGCGATCGCGCTGTCCCTGGGGATGCTGGGGGTGGCGGTCCTG 1038
              | |    |     ||||| |   |||   || |   | |   |||||  || | |    |
Db        988 CTGCTCGGACTGGCGGCCACCGCCGCCGCGTTCGCCGTGGCACTGGGCCTGTCCTGGGTC 1047

Qy       1039 GCCGGCGGTCTGCTGCACACCGTCGTCGAACAGCCCTGCATGCGCCTGTTCGGCAGCCGC 1098
                ||      || |   |   |    ||  | |    | ||||    | ||  |
Db       1048 CTGTACGAGGGCGTGGAACGCCCGGCGAGACGCCTCCTGCTCCGCCACCGCCGCTCCGCG 1107

Qy       1099 AGGTCCGCCTCCCGTCCGAAGCCCGGCGCCACCGCGGCTCCCCGGAACTCACCCGCGGCC 1158
                ||   |   |||  ||||   || |   | | | || | | ||  || |||||
Db       1108 CCACCCCGACCGACACCGCGGCCCACACCCCCGAAGCCCCACCCGGAGTCGACCACGGCC 1167

Qy       1159 GACGCGGCC 1167
                 || | |
Db       1168 CGGGCCGAC 1176


RESULT 14
US-10-496-351-23
; Sequence 23, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
;   LENGTH: 1137
;   TYPE: DNA
;   ORGANISM: Actinosynnema pretiosum
US-10-496-351-23

  Query Match             14.8%;  Score 184;  DB 12;  Length 1137;
  Best Local Similarity   50.6%;  Pred. No. 5e-35;
  Matches  501;  Conservative    0;  Mismatches  480;  Indels    9;  Gaps    2;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78
              ||  | || |||||||||| |||||  |      ||| |||    || ||||     |||
Db         31 CTTCCCTCCCTCACCGGCATCCGCGCGCCGCTGGCCCTGCTCGTGTTCGTGGCCCACGCG 90

Qy         79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCCCG 138
                |    | |   ||| || | |  |   || || | | ||    |  | ||   ||
Db         91 CTCGGCTCCGCCCGGTTCTTCGCCGACGAATCGGTCAACTCGCTCGGGTTCCTCCTGCCC 150

Qy        139 CTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCG 198
               |||  ||  |||| | ||| | |||||| | || || || ||||| || |  ||||||
Db        151 TACGGGCCCGCGGCGCTGTCGTTGTTCTTCGTGCTCAGCGGGTTCGTGCTGGTGTGGTCG 210

Qy        199 GCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCG 258
```

```
Db        211 GAGCCGTGGCGCGAGGGCGTGGGCCCCTACTTCCGACGCCGGGTCGTGCGCATCCTGCCC 270

Qy        259 CTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCT---GTCGGAGCCG 315

Db        271 ACCCACGTGCTGACGTGGGCCGCGGTGCTGCTGCTCCTGGCGGCGCTCGGCCCGCTGCCG 330

Qy        316 GTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCC 375

Db        331 CTGCTGGGGCCGCTGCCCGAGGTGGGGCCCGCGCTGGTCAACCTGTCGCTGCTCCAGTCG 390

Qy        376 TGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGT 435

Db        391 CTCGTGCCGCTGCCCGACTACCTGCTGTCGGTCAACGGCATCAACTGGTCGGTGTCCTGC 450

Qy        436 GAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCA 495

Db        451 GAGGTGGTGTTCTACCTGCTGCTGCCGCTGCTGTCCCGGCCGCTGCTGCGGGTGCCCGAC 510

Qy        496 CGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCC------C 549

Db        511 CACCGGTTGTGGGCGTGCTTCGGCGTGCTCGCCGCGGTGGTGCTCGCGCTGCCGGGCGTG 570

Qy        550 CTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAG 609

Db        571 ATCGGGGCGCTGGTGGACGGGCCGCCGTGGGCGCTGTGGCCGCCGCTGTCGTTCGAGCAG 630

Qy        610 GTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTG 669

Db        631 GCGTGGCTGGTGAACTTCTTCCCGCTGACCCGGTTGCCGGAGTTCCTGATGGGCGTGGTG 690

Qy        670 ATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCG 729

Db        691 CTCGCCAGGATCGTGGCCACCGGGCGGTGGCGGCCGGTGCGGGCGTGGTGGCCGCTGCTG 750

Qy        730 CTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCC 789

Db        751 GGGGTGGCGGCGGTGTGGGCGCTGCTGCCGGTGCTGCCGCAGGTGTACGCGCGCAGCGCG 810

Qy        790 TGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGC 849

Db        811 ATCGCGGCGGTCCCGCTGGCGCTGCTCGTCCCCGTCATCGCCGTGCGGGACCTGGAGGGG 870

Qy        850 CGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTC 909

Db        871 CGGCGGTCCGTGCTGTCCCGGCGGTGGGTGCGGGTCGCGGGGGACATGAGCTACGCGACC 930

Qy        910 TACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGC 969

Db        931 TACCTGCTGCACTGGCCGCTGCTCGCGGTCGCCAAGCACGTGGCGGGGGACCGGCTGCTC 990

Qy        970 TACGAGCGGCAGTGGAGCACCCCGGCCGCG 999

Db        991 GGGCTGGGGGAGGGGCTGCTGCTGGTGGCG 1020


RESULT 15
US-10-496-351-56/c
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
;  APPLICANT: Floss, Heinz
;  APPLICANT:  Yu, Tin-Wein
;  APPLICANT:  Leistner, Eckard
;  TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
;  TITLE OF INVENTION:  Ansamitocin
;  FILE REFERENCE: UWASH-06712
;  CURRENT APPLICATION NUMBER: US/10/496,351
;  CURRENT FILING DATE:  2004-05-19
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
;;  LENGTH: 82746
;  TYPE: DNA
;  ORGANISM: Actinosynnema pretiosum
US-10-496-351-56

  Query Match            14.8%;  Score 184;  DB 12;  Length 82746;
  Best Local Similarity  50.6%;  Pred. No. 3.6e-35;
  Matches  501;  Conservative    0;  Mismatches  480;  Indels    9;  Gaps    2;

Qy         19 CTGTCGTCACTCACCGGCCTCCGCTGGTATGCGGCGCTGACGGTATTCGCCTGTCACATC 78

Db      44758 CTTCCCTCCCTCACCGGCATCCGCGCGCCGCTGGCCCTGCTCGTGTTCGTGGCCCACGCG 44699

Qy         79 GCCCAGCAGGGCTTCTTCGCCGACCAGCAGGTGGGCAGCGCACTGCTGCACATCACCCCG 138

Db      44698 CTCGGCTCCGCCCGGTTCTTCGCCGACGAATCGGTCAACTCGCTCGGGTTCCTCCTGCCC 44639

Qy        139 CTCGGTTCCATGGCGGTCTCGATCTTCTTCATACTGAGTGGATTCGTCCTCGCCTGGTCG 198

Db      44638 TACGGGCCCGCGGCGCTGTCGTTGTTCTTCGTGCTCAGCGGGTTCGTGCTGGTGTGGTCG 44579

Qy        199 GCCCGCGACGAGGACTCCGTGCCGACTTTCTGGCGGCGCCGCATCGCGAAGATCTATCCG 258

Db      44578 GAGCCGTGGCGCGAGGGCGTGGGCCCCTACTTCCGACGCCGGGTCGTGCGCATCCTGCCC 44519

Qy        259 CTGCATCTCGCGACGTTCGGCATCGCGGCTCTCATCATTTTCTCCCT---GTCGGAGCCG 315

Db      44518 ACCCACGTGCTGACGTGGGCCGCGGTGCTGCTGCTCCTGGCGGCGCTCGGCCCGCTGCCG 44459

Qy        316 GTACTTCCCGGCGGTTCCGTATGGGACGGGCTGGTGCCCAATGTTCTGCTCGTGCAGTCC 375

Db      44458 CTGCTGGGGCCGCTGCCCGAGGTGGGGCCCGCGCTGGTCAACCTGTCGCTGCTCCAGTCG 44399

Qy        376 TGGCTTCCCGACGCGACCCTCACGGCCAGTTTCAACACGCCCAGCTGGTCGCTCTCCTGT 435

Db      44398 CTCGTGCCGCTGCCCGACTACCTGCTGTCGGTCAACGGCATCAACTGGTCGGTGTCCTGC 44339

Qy        436 GAGATCGCCTTCTATCTGTCGTTCCCGCTGTGGTACCGGCTGGTGCGCAGGATTCCCGCA 495

Db      44338 GAGGTGGTGTTCTACCTGCTGCTGCCGCTGCTGTCCCGGCCGCTGCTGCGGGTGCCCGAC 44279

Qy        496 CGGCGGCTGTGGTGGTGCGCCGCGGGGATCGCCGTGGCCGTGACGTGTGTGCC------C 549

Db      44278 CACCGGTTGTGGGCGTGCTTCGGCGTGCTCGCCGCGGTGGTGCTCGCGCTGCCGGGCGTG 44219

Qy        550 CTGCTGGCGGGCCTGCTCCCGGCGAGCGAGGAGGTGGCCCCCGGGATGTCGCTCAACGAG 609

Db      44218 ATCGGGGCGCTGGTGGACGGGCCGCCGTGGGCGCTGTGGCCGCCGCTGTCGTTCGAGCAG 44159

Qy        610 GTCTGGTTCGCGTACTGGCTTCCGCCGGTGCGCATGCTGGAGTTCGTCCTCGGCATCGTG 669

Db      44158 GCGTGGCTGGTGAACTTCTTCCCGCTGACCCGGTTGCCGGAGTTCCTGATGGGCGTGGTG 44099

Qy        670 ATGGCGCTGATCCTGCGCGCGGGGATCTGGAAGGGCCCCGGTCCGGCGGTCTGCACGGCG 729

Db      44098 CTCGCCAGGATCGTGGCCACCGGGCGGTGGCGGCCGGTGCGGGCGTGGTGGCCGCTGCTG 44039

Qy        730 CTCCTCGCCGCGAGTTACGGCCTCACCCAGATGGTGCCCCCGATCTTCACCCTCGTCGCC 789

Db      44038 GGGGTGGCGGCGGTGTGGGCGCTGCTGCCGGTGCTGCCGCAGGTGTACGCGCGCAGCGCG 43979

Qy        790 TGCTCCGTCGTACCGGCCGCGCTGCTGATCACGGCGCTGGCCGACGCCGACGTGCACGGC 849

Db      43978 ATCGCGGCGGTCCCGCTGGCGCTGCTCGTCCCCGTCATCGCCGTGCGGGACCTGGAGGGG 43919

Qy        850 CGGCGCACGGGGCTGCGTTCGGCGACGCTGGTGCGGCTGGGCCAGTGGTCCTTCGCCTTC 909

Db      43918 CGGCGGTCCGTGCTGTCCCGGCGGTGGGTGCGGGTCGCGGGGGACATGAGCTACGCGACC 43859

Qy        910 TACCTGGTCCACTTCCTGATCATCCGCTACGGACACCGGCTGATGGGCGGCGATCTGGGC 969

Db      43858 TACCTGCTGCACTGGCCGCTGCTCGCGGTCGCCAAGCACGTGGCGGGGGACCGGCTGCTC 43799

Qy        970 TACGAGCGGCAGTGGAGCACCCCGGCCGCG 999
```

```
                    | || || ||      | ||  |||
Db      43798 GGGCTGGGGGAGGGGCTGCTGCTGGTGGCG 43769
```

Search completed: August 25, 2006, 22:38:39
Job time : 2232 secs

This Page Blank (uspto)